DOCUMENT RESUME

ED 445 117 TM 031 800

AUTHOR	Murray, T. Scott; Kirsch, Irwin S.; Jenkins, Lynn B.
TITLE	Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-98-053
ISBN	ISBN-0-16-049372-2
PUB DATE	1998-01-00
NOTE	457p.
AVAILABLE FROM	U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328. Web site: http://nces.ed.gov.
PUB TYPE	Collected Works - General (020) -- Numerical/Quantitative Data (110) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC19 Plus Postage.
DESCRIPTORS	*Adult Literacy; *Adults; Foreign Countries; *International Studies; *Research Methodology; Surveys; Tables (Data)
IDENTIFIERS	Canada; Europe; *Organisation for Economic Cooperation Development; United States

ABSTRACT

In December 1995, the Organisation for Economic Co-Operation and Development (OECD) and Statistics Canada jointly published the results of the first International Adult Literacy Survey (IALS). For this survey, representative samples of adults aged 16 to 65 were interviewed and tested in their homes in Canada, France, Germany, the Netherlands, Poland, Sweden, Switzerland, and the United States. This report describes how the survey was conducted in each country and presents all available evidence on the extent of bias in each country's data. Potential sources of bias, including sampling error, non-sampling error, and the cultural appropriateness and construct validity of the assessment instruments, are discussed. The chapters are; (1) "Introduction" (Irwin S. Kirsch and T. Scott Murray); (2) "Sample Design" (Nancy Darcovich); (3) "Survey Response and Weighting" (Nancy Darcovich); (4) "Non-Response Bias" (Nancy Darcovich, Marilyn Binkley, Jon Cohen, Mats Myrberg, and Stefan Persson); (5) "Data Collection and Processing" (Nancy Darcovich and T. Scott Murray); (6) "Incentives and the Motivation To Perform Well" (Stan Jones); (7) "The Measurement of Adult Literacy" (Irwin S. Kirsch, Ann Jungeblut, and Peter B. Mosenthal); (8) "Validity Generalization of the Assessment across Countries" (Don Rock); (9) "An Analysis of Items with Different Parameters across Countries" (Marilyn R. Binkley and Jean R. Pignal); (10) "Scaling and Scale Linking" (Kentaro Yamamoto); (11) "Proficiency Estimation" (Kentaro Yamamoto and Irwin S. Kirsch); (12) "Plausibility of Proficiency Estimates" (Richard Shillington); and (13) "Nested-Factor Models for the Swedish IALS Data" (Bo Palaszewski). Fourteen appendixes contain supplemental information, some survey questionnaires, and additional documentation for various chapters. (Contains 94 tables, 12 figures, and 74 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

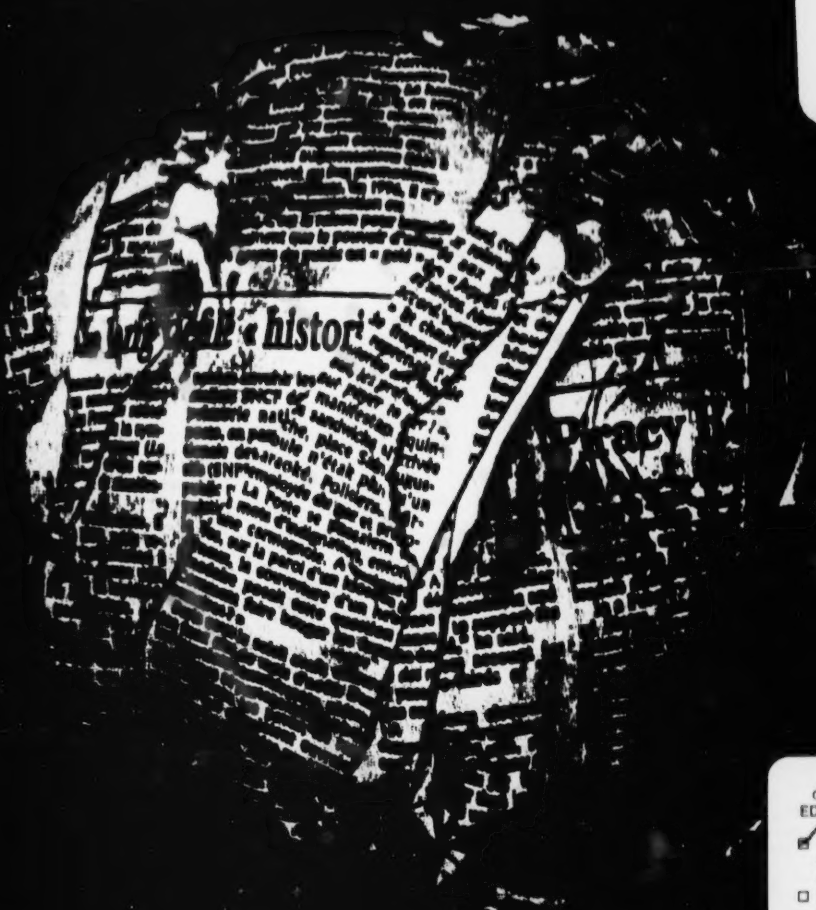
ED 445 117

NATIONAL CENTER FOR EDUCATION STATISTICS

Adult Literacy in OECD Countries

Technical Report on the First International Adult Literacy Survey

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view. CE



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM031800

U.S. Department of Education
Office of Educational Research and Improvement
NCES 98-053

NATIONAL CENTER FOR EDUCATION STATISTICS

Adult Literacy in OECD Countries

Technical Report on the First International Adult Literacy Survey

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: IM
In our judgment, this document is also of interest to the Clearinghouses noted to the right. Indexing should reflect their special points of view CE

T. Scott Murray
Statistics Canada

Irwin S. Kirsch
Educational Testing Service

Lynn B. Jenkins

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement
NCES 98-053

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208–5574

January 1998

The NCES World Wide Web Home Page is
http://nces.ed.gov

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*, NCES 98–053, by T. Scott Murray, Irwin S. Kirsch, and Lynn B. Jenkins. Project Officer, Marilyn Binkley. Washington, DC: 1998

Content Contact:
Marilyn Binkley
202–219–2195

Contents

Preface 11

Chapter 1: **Introduction**—*Irwin S. Kirsch and T. Scott Murray* 13

1.1 Survey Goals, Definitions, and Methods 13

The International Adult Literacy Survey: A brief history 14

Defining and measuring literacy 17

Conducting the survey 18

Reporting the results 20

1.2 Objectives of This Report 20

1.3 A Framework for Evaluating Data Quality 21

1.4 Organization of This Report 22

Part I: **Sample Design Error and Non-response Bias** 23

Chapter 2: **Sample Design**—*Nancy Darcovich* 25

2.1 Introduction 25

2.2 Target Population and Frame Coverage 26

2.3 Sample Design 29

2.4 Concerns About Non-probability-based Sampling 36

2.5 Conclusions 39

Chapter 3: **Survey Response and Weighting**—*Nancy Darcovich* 41

3.1 Introduction 41

3.2 Survey Response 42

3.3 Weighting 46

3.4 Benchmarking 50

3.5 Internal and External Checks 51

3.6 Sampling Error 53

3.7 Conclusions 53

Contents

Chapter 4: **Non-response Bias**—*Nancy Darcovich, Marilyn Binkley, Jon Cohen, Mats Myrberg, and Stefan Persson* ... 55
4.1 Introduction ... 55
4.2 Canadian Study of Non-response Bias ... 55
4.3 United States Study of Non-response Bias ... 61
4.4 Swedish Study of Non-response Bias ... 67
4.5 Conclusions ... 71

Part II: **Non-sampling Error** ... 73

Chapter 5: **Data Collection and Processing**—*Nancy Darcovich and T. Scott Murray* ... 75
5.1 Introduction ... 75
5.2 Model Procedures Manuals and Instruments ... 76
  Background questions ... 76
  Literacy tasks ... 76
  Standardized non-response coding ... 77
5.3 Field Operations ... 77
5.4 Scoring ... 83
  Analysis of pilot survey data ... 83
  Quality control ... 86
  Intra-country rescoring ... 87
  Inter-country rescoring ... 88
5.5 Data Capture, Data Processing, and Coding ... 89
5.6 Conclusions ... 90

Chapter 6: **Incentives and the Motivation to Perform Well**—*Stan Jones* ... 93
6.1 Introduction ... 93
6.2 Participation Bias ... 94
6.3 Motivation to Perform Well ... 96
6.4 Omit Rate ... 96
6.5 Not Reached Rate ... 97
6.6 Time Taken on Test ... 98
6.7 Conclusions ... 99

Contents

Part III: **Construct Validity and Proficiency Estimation** 101

Chapter 7: **The Measurement of Adult Literacy**—*Irwin S. Kirsch, Ann Jungeblut, and Peter B. Mosenthal* 105

7.1 Introduction 105
7.2 The Young Adult Literacy Survey 106
Understanding task difficulty 111
7.3 The U.S. Department of Labor Literacy Survey 113
Prose literacy 113
Document literacy 118
Quantitative literacy 122
Establishing literacy levels 125
7.4 The National Adult Literacy Survey 129
7.5 Conclusions 134

Chapter 8: **Validity Generalization of the Assessment Across Countries**—*Don Rock* 135

8.1 Introduction 135
8.2 Sample and Methods 136
8.3 Results 137
8.4 Conclusions 142

Chapter 9: **An Analysis of Items with Different Parameters Across Countries**—*Marilyn R. Binkley and Jean R. Pignal* 143

9.1 Introduction 143
9.2 A Theory of the Cognitive Processes Underlying Literacy Proficiency 146
9.3 Application of the Theory to the IALS 146
9.4 Analysis of Items 147
Example 1 147
Example 2 149
Examples 3 and 4 150
Examples 5 and 6 152
Example 7 154
9.5 Do the Examples Fit the Theory? 158
9.6 Conclusions 160

BEST COPY AVAILABLE

Contents

Chapter 10: **Scaling and Scale Linking**—*Kentaro Yamamoto* 161
- 10.1 Introduction 161
- 10.2 Analysis of Items Using Mantel-Haenszel Statistics 162
- 10.3 Block Order Effects 164
- 10.4 Scaling Model 166
- 10.5 Item Parameter Estimation 167
- 10.6 Scale Linking Procedures 171
- 10.7 Evaluation of Differential Group Performance 175
- 10.8 Conclusions 177

Chapter 11: **Proficiency Estimation**—*Kentaro Yamamoto and Irwin S. Kirsch* 179
- 11.1 Introduction 179
- 11.2 Plausible Values and Proficiency Estimation 179
- 11.3 Proficiency Estimation in the IALS 180
- 11.4 Classical Test Statistics and Proficiency Values 183
  - Proportion of correct responses 183
  - Non-response 183
  - Omitted responses 184
  - Not-reached responses 185
- 11.5 Missing Cognitive Data 187
- 11.6 Imputing the Literacy Proficiencies of Non-respondents 188
- 11.7 Conclusions 190

Chapter 12: **Plausibility of Proficiency Estimates**—*Richard Shillington* 193
- 12.1 Introduction 193
- 12.2 Methodology 193
- 12.3 Logistic Regressions 196
- 12.4 Assessing the Multiple Correlation Between Literacy and Demographic Characteristics 197
- 12.5 Conclusions 201

Chapter 13: **Nested-factor Models for the Swedish IALS Data**—*Bo Palaszewski* 203
- 13.1 Introduction 203
- 13.2 Data 204
- 13.3 Data Analysis 205
- 13.4 Conclusions 208

References 209

Contents

Appendices 215
Appendix A: Review of Methodology
Appendix B: IALS Background Questionnaire
Appendix C: Model Procedures Manual
Appendix D: Model Interviewer's Manuals
Appendix E: Sampling Guidelines
Appendix F: International Data Structure
Appendix G: Example of a Scoring Rubric
Appendix H: Example of a Coding Scheme
Appendix I: Additional Documentation – Chapter 4
Appendix J: Additional Documentation – Chapter 5
Appendix K: Additional Documentation – Chapter 10
Appendix L: Additional Documentation – Chapter 11
Appendix M: Additional Documentation – Chapter 12
Appendix N: Additional Documentation – Chapter 13

About the Authors

Marilyn Binkley, National Center for Education Statistics, Washington, D.C., U.S.A.

Jon Cohen, Pelavin Research Center, American Institutes for Research, Washington, D.C., U.S.A.

Nancy Darcovich, Statistics Canada, Ottawa, Ontario, Canada

Lynn B. Jenkins, Northford, Connecticut, U.S.A.

Stan Jones, Carleton University, Ottawa, Ontario, Canada

Ann Jungeblut, Educational Testing Service, Princeton, New Jersey, U.S.A.

Graham Kalton, WESTAT, Rockville, U.S.A.

Irwin S. Kirsch, Educational Testing Service, Princeton, New Jersey, U.S.A.

Lars Lyberg, Statistics Sweden, Stockholm, Sweden

Peter B. Mosenthal, Syracuse University, Syracuse, New York, U.S.A.

T. Scott Murray, Statistics Canada, Ottawa, Ontario, Canada

Mats Myrberg, University of Linköping, Sweden

Bo Palaszewski, University of Göteborg, Sweden

Stefan Persson, National Agency for Education, Stockholm, Sweden

Jean R. Pignal, Statistics Canada, Ottawa, Ontario, Canada

Jean-Michel Rempp, INSEE, Paris, France

Don Rock, Educational Testing Service, Princeton, New Jersey, U.S.A.

Richard Shillington, Statistics Canada (consultant), Ottawa, Ontario, Canada

Albert Tuijnman, Organisation for Economic Co-operation and Development, Paris, France

Kentaro Yamamoto, Educational Testing Service, Princeton, New Jersey, U.S.A.

Preface

In December 1995 the Organisation for Economic Co-operation and Development (OECD) and Statistics Canada jointly published the results of the first International Adult Literacy Survey (IALS). For this survey, conducted during the autumn of 1994, representative samples of adults aged 16 to 65 were interviewed and tested in their homes in Canada (English and French-speaking provinces), France, Germany, the Netherlands, Poland, Sweden, Switzerland (French and German-speaking cantons), and the United States.

The IALS established an important precedent by extending educational measurement and assessment to the adult population. The results enable us for the first time to profile the literacy and numeracy of adults in different countries as well as to make international comparisons of particular subpopulations of interest—for example, adults in the labor force.

For obvious reasons, a survey such as this immediately prompts questions about ranking. Which country came in first, and which came in last? Ranking the countries based on relevant aspects of literacy proficiency can be useful in establishing benchmarks, setting standards, and evaluating literacy. Yet this was not the sole or even the most important reason why the participating countries agreed to field the assessment. Rather, they expected the survey to yield insights into the mix of factors that influence the development of adult literacy, regardless of nationality. Cross-national comparisons would add contextual information needed to interpret the survey results.

Whether the IALS meets the high expectations of the participating countries and of other interested observers depends heavily on the reliability, validity, and comparability of the survey data. Given the importance placed on a study such as this, it was imperative that high standards of data quality be achieved by all participating countries.

This report describes how the survey was conducted in each country and presents all available evidence on the extent of bias in each country's data. Potential sources of bias—including sampling error, non-sampling error, and the cultural appropriateness and construct validity of the assessment instruments—are discussed. This evidence will enable readers to draw their own informed conclusions about the quality of the data and value of the study.

As is the case in any household survey, the IALS results contain a degree of error, although the standard errors of the estimates are typically quite small. The estimates are also subject to bias, the precise magnitude and direction of which are unknown. Consider, for example, that an estimated 24 percent of United States adults scored at the lowest level of document literacy, while the corresponding figures are 9 percent for Germany and 36 percent for France. The true figures may be somewhat higher or lower, depending on the extent and direction of bias in the data for each country.

Even assuming maximum values for error and bias, however, the fundamental conclusion that emerges from the IALS results remains unchanged: millions of adults in each country face tremendous difficulties in coping with the literacy demands placed upon them in today's complex and changing economy and society.

Albert Tuijnman

Education and Training Division

Organisation for Economic Co-operation and Development

BEST COPY AVAILABLE

Chapter 1:

Introduction

Irwin S. Kirsch and T. Scott Murray

1.1 Survey Goals, Definitions, and Methods

In recent years, adult literacy has come to be seen as crucial to the economic performance of industrialized nations. Literacy is no longer defined merely in terms of a basic threshold of reading ability which everyone growing up in developed countries is expected to attain. Rather, literacy is now equated with an individual's ability to use written information to function in society. Unlike their predecessors, adults today need a higher level of literacy to function well, because society has become more complex and low-skill jobs are disappearing. Inadequate levels of literacy in a broad section of the population may therefore have serious implications, even threatening a nation's economic strength and social cohesion.

Because of these high stakes, governments have a growing interest in understanding the level and distribution of literacy within their adult populations, and in learning what can be done to improve literacy. Accordingly, in recent years, many governments have tried for the first time to measure adult literacy directly. Pioneering studies (Kirsch and Jungeblut 1986; Kirsch and Mosenthal 1990; Statistics Canada 1991; Kirsch, Jungeblut, and Campbell 1992; Kirsch, Jungeblut, Jenkins, and Kolstad 1993) published in North America in the early 1990s revealed that significant percentages of adults lacked the literacy skills they were likely to need in everyday life. In 1992, the Organisation for Economic Co-operation and Development (OECD) concluded that low literacy levels were a serious threat to economic performance and social cohesion (OECD 1992). Yet a lack of comparable international data prevented a broader inquiry into literacy problems and consequent policy lessons across industrialized countries.

The International Adult Literacy Survey (IALS) was undertaken by nine governments and three intergovernmental organizations in a collaborative effort to fill this need for information. In this survey, large samples of adults (ranging from

1,500 to 8,000 per country) in Europe and North America were given the same broad test of their literacy skills during the autumn of 1994. The results provide the most detailed portrait ever created on the condition of adult literacy and its relationship with an array of background and demographic characteristics. The study's findings were summarized in a report published in December 1995, entitled *Literacy, Economy and Society: Results of the first International Adult Literacy Survey*.

The International Adult Literacy Survey: A brief history

Studies such as the International Adult Literacy Survey do not arise spontaneously but represent one step in an evolving research tradition (Kirsch and Jungeblut 1986; Wickert 1989). IALS represented a first attempt to assess the skills of entire adult populations in a comparative framework, building on advances in theory and practice in the fields of educational assessment, reading theory, psychometrics and sample survey methods. Thus, it is difficult, if not impossible, to appreciate the study's strengths and weaknesses without an appreciation of the events which precipitated the study, what assumptions were made in planning the study and how the study unfolded in time.

From a scientific viewpoint, IALS can trace it's roots to the 1985 Young Adult Literacy Survey (YALS). Conducted by the Educational Testing Service on behalf of the United States Department of Education, YALS was the first large scale attempt to employ the techniques of educational assessment in a household survey environment. The YALS proficiency measures were based on a new conception of reading literacy developed by Irwin Kirsch of the Educational Testing Service and Peter Mosenthal of Syracuse University, a framework which explained an unprecedented amount of the observed variance in proficiency. The study was also the first to employ a complex matrix design and related advances in psychometrics to estimate population level proficiencies.

The YALS study was followed by a study of job seekers sponsored by the United States Department of Labor (the DOL study) (Kirsch, Jungeblut, and Campbell 1992), and ultimately by the National Adult Literacy Survey (NALS) (Kirsch, Jungeblut, Jenkins, and Kolstad 1993), a mammoth assessment of the entire adult population. Each successive study allowed the theory and procedures to be further refined. In order to maximise the comparability of estimates across countries, the IALS study chose to adopt the NALS methodology and scales. An overview of the NALS methods and results may be found in the NALS technical report (United States Department of Education, in progress).

In 1989, Statistics Canada fielded a national assessment of adult literacy called the Survey of Literacy Skills Used in Daily Activities (LSUDA) (Statistics Canada 1991). Although LSUDA used a less elaborate assessment design than the prototype United States studies cited above, the study was the first to attempt to assess literacy in a valid and reliable way across language and culture, in this case English and French. Thus, LSUDA was the first study to empirically demonstrate that the Kirsch-Mosenthal framework could support such measures. LSUDA was also the first to report results in proficiency levels, an innovation adopted by the NALS study.

From a policy perspective, the North American studies cited above revealed a strong relationship between literacy and economic and opportunity and success and large numbers of adults with low literacy skills. While concerned about these findings policy makers were left without any way to judge their possible impact on economic competitiveness. By 1990, this concern had translated into a strong desire to obtain comparable data from other countries, preferably from key trading partners. The background questionnaire which was proposed to participants was designed to reveal relationship of literacy to a range of key demographic variables and economic outcomes, such as income, weeks worked and weeks unemployed.

At this stage Statistics Canada and ETS joined forces with the Organisation for Economic Co-operation and Development in Paris, the UNESCO Institute for Education in Hamburg and the Commission of the European Communities to recruit national governments into the study. Key elements of the study objectives and design were set out in the document *An International Assessment of Adult Literacy: A Proposal* (Statistics Canada 1990). Nine countries eventually agreed to participate in the development and validation of the assessment instruments, and, after considerable negotiation, to contribute to the international overheads associated with the project. It must be noted, however, that Canada and the United States ultimately absorbed the lions share of these international overheads. Countries were reluctant to contribute more towards these overheads because of the relatively high cost of simply fielding the study and the fact that IALS represented an unknown quantity—as the first of its kind—so they were reluctant to invest large sums of money on something which had yet to demonstrate its relevance for policy.

From the beginning IALS success rested on the ability of the study to meet the following conditions:

First, each country would have to select a scientifically credible sample of the adult population.

Second, each country would have to adapt the background questionnaire and assessment instruments in a way which maintained their theoretic and conceptual integrity.

Third, countries would have to administer the survey instruments in keeping with a set of prescribed collection procedures.

Finally, countries would have to edit their data files and submit clean files in a standardised format.

While all of the forgoing elements were deemed to be important, the assumption going into the study was that instrument adaptation, particularly of the assessment instruments, was the design element which carried the most inherent risk. Thus, although quality assurance procedures were put in place for each element, most of the available resources were directed to the assessment and psychometric aspects of the study. In retrospect this assumption proved to be incorrect. The evidence presented in this volume and our own professional judgement suggests that the psychometric aspects of the study actually performed far better than expected. For example, the assessment was engineered to withstand a loss of up to 40% of items before a common scale would be lost. Actual losses, where items were either dropped or assigned idiosyncratic item parameters, were on the order of 10%. Again, in retrospect, more attention should have been paid to quality assurance of other design elements.

With the exception of Sweden and Ireland, who joined the study late, all countries participating in the first round of IALS data collection also participated in the development of the background questionnaire and the assessment instruments. All countries fielded a pilot study which demonstrated, among other things, that the assessment was performing as expected.

All nine countries proceeded to field the main assessment in 1994. Although data collection ended more or less on schedule, most countries seriously underestimated the amount of effort required to edit and score the data file. Thus, only Canada was able to make the agreed to deadline for submission of clean data files to ETS for scaling. Due to the untimely death of its national study manager, Ireland was not able to provide a clean data set until early 1996. Proficiency estimates for eight countries became available in July, 1995. At this point, despite having participated in every phase of development, the Government of France began to express misgivings about the quality of the study. In order to address these concerns in a neutral way, Statistics Canada engaged the services of Messrs. Rempp, Lyberg and Kalton to review all aspects of the study. Their report, included herewith as Appendix A, identified several weaknesses in the study but recommended that publication should proceed as planned. Despite this recommendation the Government of France chose to withdraw from the study. Data for the remaining seven countries was subsequently published in *Literacy, Economy and Society: Results of the first International Adult Literacy Survey* in December, 1995.

This report has three objectives. First, the report attempts to document how the survey was conducted in each country and what was known at the time of the release of *Literacy, Economy and Society* about the extent of bias in each countries data. Second, the report includes the results of additional analyses which were undertaken to shed empirical light on the French concerns. These analyses are presented in the Appendices of the report. Thus, this volume should provide readers to draw their own informed conclusions about the quality of the data and the value of the study. Finally, the document serves the dual purpose of sensitising countries considering participation in future rounds of IALS data collection about the need to adhere to the stated design specifications, and of informing the introduction of enhanced quality assurance procedures into the survey methodology. These enhancements will be documented in future editions of this report.

As noted above, the IALS study represents an important step in the evolution of a relatively new research tradition: the assessment of the performance of entire adult populations in a comparative framework. While far from perfect, the IALS study has provided new insight into the causes and consequences of literacy skill in advanced economies, insight which has already had a major impact on policy in both Canada and Sweden. Data from a second round of data collection, involving Australia, New Zealand, Great Britain, Northern Ireland and Flemish Belgium is scheduled for release in October, 1997 and planning for a third round of collection, involving an additional ten countries, is well advanced. Each of these new rounds of collection have benefited greatly from the refined quality assurance procedures which the evidence presented in this volume has enabled.

The IALS venture was initiated with two fundamental goals. The first objective was to develop scales that would permit comparisons of the literacy performance of adults with a wide range of abilities. Then, if such an assessment could be created, the second goal was to describe and compare the demonstrated literacy skills of adults in different countries. This second objective presented the challenge of comparing literacy across cultures and across languages.

Statistics Canada, the statistical arm of the Canadian government, and Educational Testing Service, the leading private testing organization in the United States, coordinated the development and management of the IALS. These organizations were guided by national research teams from the participating countries, which assisted in developing the survey design. The IALS built on earlier national surveys that had been carried out in the United States and Canada to assess adult literacy. It also drew upon recent research and on methodological and technical advances in assessment and psychometrics.

The IALS has provided a rich array of data on the literacy skills of adults in seven advanced industrialized nations—data that are comparable across cultures and languages.[1] The survey results have generated wide interest and will be useful for many different kinds of analysis. From the scientific perspective, the empirical validation of the literacy models underlying the assessment may be the study's most important finding.

The remainder of this introduction summarizes how literacy was defined and measured in the IALS, how the survey was conducted, how the data were reported, and how data quality is evaluated.

Defining and measuring literacy

Many studies have treated literacy as a condition that adults either have or do not have, and thereby tried to count the number of illiterate members of the population. Such efforts typically define literacy in terms of the number of years of schooling completed, or by grade-level scores on school-based reading tests.

The IALS survey design team agreed that it would be undesirable to establish a single international standard for literacy. Such a standard would not only be arbitrary, but would also fail to acknowledge the multifaceted nature of literacy and the complexity of the literacy problem. Therefore, the participating countries concurred that, in common with recent North American and Australian surveys (The Commonwealth Department of Employment, Education and Training 1989), the IALS would define literacy in terms of a mode of adult behavior, namely:

> *Using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.*

This definition attempts to encompass a broad set of information-processing skills that adults may use in performing different types of tasks at work, at home, or in their communities. Some other types of knowledge and skill (including teamwork, interpersonal skills, and other communication skills) were also recognized as being important but could not be measured with the resources available.

According to the IALS definition, literacy is neither a single skill used in dealing with all types of text, nor an infinite set of skills, each particular to a different type of material. Thus, following the example of the North American studies noted earlier, the IALS team defined three domains of literacy:

Prose literacy—the knowledge and skills needed to understand and use information from texts including editorials, news stories, poems, and fiction;

Document literacy—the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables, and graphics; and

Quantitative literacy—the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a checkbook, calculating a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

Rather than define a threshold for competency—a standard that distinguishes the so-called "literate" from the "illiterate"—researchers constructed a scale from 0 to 500 in each of these three literacy domains. Tasks of varying difficulty can be placed along these scales. A person's literacy ability in each domain is expressed by a score, which is the point on the scale at which he or she has an 80 percent chance of successfully performing a given literacy task. Individuals can then be grouped into five levels of literacy, defined by score ranges (Level 1 includes scores from 0 to 225, Level 2 contains scores from 226 to 275, and so on). These levels are useful in analyzing and reporting the survey results and in designing remedial programs.

---

[1] Ireland was unable to complete its data processing in time to be included in the *Literacy, Economy and Society* report, and France withdrew from the study prior to the report's publication.

The use of three parallel literacy scales makes it possible to profile and compare the various types and levels of literacy demonstrated by adults in different countries and by subgroups within those countries. The scales also help policy makers, business leaders, educators, and others to understand the broad and diverse nature of literacy.

Conducting the survey

Each country that participated in the IALS agreed to draw a probability sample that would accurately represent its civilian, non-institutionalized population aged 16 to 65 (Table 1.1). Readers will note that these sample sizes vary considerably from country to country. The IALS design document stipulated that countries field a sample sufficiently large to yield 3,000 completed cases. In many cases countries were unable or unwilling to commit sufficient resources to meet this requirement, a fact which constrains the level of analysis which may be undertaken. In all cases, however, sample sizes were sufficiently large to support the estimation of reliable IRT item parameters. Several countries also included older adults in their samples. In six countries, the survey was carried out in the national language. In Canada, respondents were given a choice of answering questions in either English or French. In Switzerland, individuals in French-speaking and German-speaking cantons were required to respond in those respective languages (Italian-and Rhaeto-Romanic-speaking regions were excluded). When a respondent could not speak the designated language, the interviewer still attempted to obtain answers to the background questions so that the respondent's literacy levels could be estimated; this would reduce the possibility of biased results.

Table 1.1: **Survey coverage, language of survey, and sample yields**

Country	Population aged 16 to 65 covered by the survey	Language of survey	Sample yield
Canada	13,676,612 4,773,648	English French	3,130 1,370
France	36,432,474	French	2,996
Germany	53,826,289	German	2,062
Ireland	2,174,000	English	2,423
Netherlands	10,460,359	Dutch	2,837
Poland	24,475,649	Polish	3,000
Sweden	5,361,942	Swedish	2,645
Switzerland	1,008,275 3,144,912	French German	1,435 1,393
United States	161,121,972	English	3,053

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

A direct survey of adult literacy necessitates visits to people's homes so that their abilities can be assessed in a manner usually done in schools. Thus, the IALS, like the two preceding national studies in North America, was unusual because it combined the techniques of household-based survey research with those of educational testing.

IALS respondents were asked to perform a diverse set of literacy tasks and to complete a background questionnaire which gathered detailed information about their demographic and other characteristics. These data provide a means for exploring how literacy is connected to social, educational, economic, and other variables, and for exploring whether or to what extent these relationships are similar across cultures.

In surveying literacy internationally, it was necessary to ensure that the assessment tasks reflected the various languages and cultures of participating countries. Study managers from each participating country were therefore encouraged to submit materials such as news articles and documents that could be used to create such tasks. The goal was to build a new pool of literacy tasks that could be linked to established scales.

To guard against cultural bias in the results, the IALS team decided to field test a large number of tasks—considerably more than were needed to obtain statistically valid estimates of each person's literacy level. This made it possible to "cross-check" the results. Of the 175 literacy tasks that were field tested, 114 were found to be valid across cultures, and these were selected for the main assessment. Approximately half of these tasks were based on materials from outside North America. Table 1.2 indicates the sources of the material upon which the tasks were based. All of the literacy tasks were open-ended rather than multiple-choice because it was thought that adults would be more interested in performing such tasks.

Table 1.2: **Sources of texts for literacy tasks**

Country	Number of texts
Canada	2
France	2
Germany	4
Ireland	1
Mexico	2
Netherlands	9
United States	17
Other	2

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

No individual could be expected to respond to the entire set of 114 literacy tasks. Accordingly, the survey was designed to give each participant a set of tasks that was carefully selected from the total pool. At the same time, it was necessary to ensure that each task was administered to nationally representative samples of adults. This was accomplished by compiling the literacy tasks into seven blocks (labeled A through G), which in turn were compiled into seven test booklets (numbered 1 through 7), as shown in Table 1.3. Each booklet contained three blocks of tasks and was designed to take about 45 minutes to complete.

Respondents began the cognitive part of the assessment by performing a set of six "core" tasks. To avoid embarrassing adults with very low literacy skills by giving them the full assessment, only those who were able to perform at least two of the six core tasks correctly (93 percent of respondents) were given the full assessment.

Table 1.3: **Booklet configuration**

Booklet number	First block	Second block	Third block
1	A	B	D
2	B	C	E
3	C	D	F
4	D	E	G
5	E	F	A
6	F	G	B
7	G	A	C

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Reporting the results

In December 1995, Statistics Canada and OECD jointly published the IALS results in a report entitled *Literacy, Economy and Society* which was commended by ministers attending a January 1996 meeting of the OECD Education Committee. The ministers asked the survey organizers to continue the program and to identify policy options for addressing the serious literacy deficit revealed by the survey results (OECD 1996). Four countries—Australia, New Zealand, the United Kingdom, and Flemish Belgium—subsequently collected data using the IALS instruments. Data for these countries is scheduled for publication in October, 1997 in *Literacy Skills for the Knowledge Society: Further Results of the International Adult Literacy Survey*. An additional 12 countries are likely to participate in an OECD-sponsored round of data collection launched in Paris on 26-27 June 1996.

1.2 Objectives of This Report

The International Adult Literacy Survey (IALS) is the first large-scale comparative assessment of adult literacy skills undertaken to date. Because strong relationships are known to exist between literacy and economic success at both the individual and national levels, the *Literacy, Economy and Society* report was noticed in the media and by senior levels of government in many countries. The report describes significant differences in literacy profiles across participating countries—differences that may be at odds with prevailing national views or expectations. Such controversial findings have only served to heighten interest in the study.

Although the IALS appears to have been a remarkable success, some have voiced concerns about the survey and the quality of the data it has provided. At the April 1996 meeting of the OECD Education Committee, the delegate from France's Ministry of National Education stated that, when the international results became available for scrutiny in August 1995, French authorities began to question the comparability of the results across countries. In particular, they questioned the appropriateness of the assessment instruments, the validity of the sampling procedures, and the reliability of the population estimates. (France joined IALS at the earliest stage, however, and its experts collaborated with representatives from the other countries in designing the assessment instruments and conducting the field test and main data collection.)

In response to these concerns, Statistics Canada commissioned three internationally recognized statisticians to prepare a report reviewing the quality of the IALS data (Appendix A). Although these statisticians unanimously recommended publication of the study results, French authorities decided in October 1995 to withdraw their country's results from the *Literacy, Economy and Society* report. In an effort to clarify issues pertaining to data quality and to improve the collaborative process, the OECD Education Committee decided to convene a meeting of experts to discuss issues of cultural appropriateness, reliability, and validity. This report was prepared for that meeting, held at OECD Headquarters on 6 September 1996.

Given the interest generated by the IALS results as well as the concerns about data quality voiced by French authorities, it is clearly important to examine the strengths and weaknesses of the study and to investigate and document any known bias in the measures. This document is intended to fulfill these objectives by providing a framework for examining the quality of the IALS data, offering a context for interpreting the survey results, and indicating the most promising avenues for further investigation. The contents offer a comprehensive summary of IALS design, implementation, and results, and will provide useful input for subsequent survey cycles. In fact, quality assurance procedures were already significantly improved for the subsequent round of data collection in Australia, New Zealand, the United Kingdom, and the Flemish community of Belgium. Further improvements are planned for the next large-scale data collection effort, scheduled for the fall of 1997.

1.3 A Framework for Evaluating Data Quality

To lay the groundwork for this report, it is important to make explicit two key assumptions. First, the quality and usefulness of any data cannot be determined in the abstract but must be evaluated in terms of its intended uses or applications. Social scientists have an ethical duty to measure and document data quality and to communicate this information to data users within the context of their intended data use.

Although this report may help readers to judge the fitness of the IALS data for a variety of purposes, its primary intent is to confirm the study's ability to support comparisons of literacy proficiency means and distributions across countries. This caveat must itself be tempered, however, since one would be concerned only about those areas in which error is large enough to distort the study's key policy implications. Virtually all of the data presented in this volume suggest that the key findings of the IALS remain uncontested.

The second key assumption is that error is a fact of life in the social sciences. In survey research, it is common practice to distinguish between two broad categories of error: *sampling error and non-sampling error*. Because the IALS employed probability sampling, the results are subject to *sampling error*. Although small, this error is rather higher in IALS than in most studies because the cost of surveying adults in their homes is so high. Most countries simply could not afford large sample sizes.

The survey data are also subject to *non-sampling error* from several sources. One key source of this type of error is *differential coverage* across countries. Another is *non-response bias*, which occurs when different groups of sampled individuals (for example, adults in different countries, or certain groups within a country) fail to participate in the survey. Other potential sources of non-sampling error include deviations from prescribed *data collection procedures*, and *errors of logic* which result from mapping idiosyncratic national data into a rigid international

BEST COPY AVAILABLE

format. *Scoring error*, associated with scoring open-ended tasks reliably within and between countries, must also be considered. Finally, because the IALS data were collected and processed independently by the various countries, the study must cope with uneven levels of commonplace *data capture, data processing and coding errors.*

In probing the quality of the IALS data, it is important to examine both sampling and non-sampling error, but this examination is insufficient by itself. The issue of *construct validity* must also be considered. Two fundamental assumptions underlie the IALS: that an international literacy construct exists, and that this construct can be measured in a stable and reliable way across national and linguistic boundaries. To the extent that these assumptions are untrue—that is, to the extent that real linguistic differences exist among countries or that the measurement protocol does not discriminate comparably across populations—the interpretation of international differences must be undertaken with care.

When IALS was being designed, it was believed that construct validity was the major concern inherent in the study. Popular wisdom suggested that achieving psychometric equivalence across languages and cultures was impossible. As a result, much of the available data quality assurance resources were devoted to this aspect of the study. Relatively little time was spent controlling the more "pedestrian" matters of sampling and data collection, under the belief that all partners were capable of fielding the survey according to the specified design. Having established the psychometric validity of the survey and instruments in the first round of data collection, it is now necessary to refine the other data quality-related aspects of the study in future survey cycles.

1.4 Organization of This Report

Each of the following chapters addresses a distinct source of error. The sections within each chapter describe quality assurance procedures enacted to contain error from each source, and provide information about known deviations from prescribed procedures or known problems. The conclusions at the end of each chapter summarize what is known about the error from each source. In an effort to improve the report's readability, much of the detailed evidence concerning data quality has been placed in voluminous appendices.

Questions about this report or its content should be directed to:

T. Scott Murray or

International Study Director for the IALS
Special Surveys Division
Statistics Canada
5-B2, Jean Talon Building, Tunney's Pasture
Ottawa, Ontario, Canada K1A 0T6
Tel.: (613) 951-9476
Fax: (613) 951-0085
Internet: scotmur@statcan.ca

Irwin S. Kirsch

Educational Testing Service
Rosedale Road
Princeton, New Jersey 08541
United States of America
Tel.: (609) 734-1516
Fax: (609) 734-1309
Internet: ikirsch@ets.org

Part I:

Sample Design Error and Non-response Bias

Introduction to Part I

As noted earlier, the primary aims of the International Adult Literacy Survey (IALS) were to profile adult literacy within the participating countries and to compare literacy across these countries. To achieve these goals, it was necessary to ensure that the survey samples were representative of the target populations.

All of the IALS countries had to adhere to certain standards in designing and implementing the survey in order to limit the level of sample design error in the survey data. Chapter 2 documents various aspects of national sample designs and analyzes the likely impact of known deviations from the established standards. Chapter 3 discusses survey response, non-response, and weighting, and Chapter 4 discusses special studies undertaken to investigate the magnitude and nature of non-response bias in the Canadian, United States, and Swedish IALS results.

Chapter 2:

Sample Design

Nancy Darcovich

2.1 Introduction

Each country participating in IALS agreed to design a sample that would be representative of its national population aged 16 to 65 years. Explicit sample design criteria were provided, and these were extensively discussed at meetings held before the pilot and main surveys. The criteria specified that each country would need to draw a high-quality probability sample representing the adult non-institutional civilian population aged 16 to 65. Only a small percentage of exclusions was acceptable.

Initially, the specifications called for each country to draw a sample large enough to yield 3,000 completed cases, but this proved to be unrealistic. Several countries were unable to secure sufficient financial resources to support such an expensive undertaking. Accordingly, the final sample design criteria specified that each country's sample should result in at least 1,000 respondents, the minimum sample size needed to produce reliable literacy proficiency estimates. All of the IALS countries achieved the minimum sample size.

The IALS sample design criteria and their importance to the integrity of the study were emphasized repeatedly at meetings of national study directors held in Ottawa, New York, Hamburg, and Amsterdam. At the Hamburg meeting, ETS used the pilot survey results to illustrate the measurement protocol's sensitivity to the representativity of the sample and other subtle influences. This theme was reiterated at the Amsterdam meeting, where participants held a detailed discussion about non-response coding and its relationship to ability estimation. It is therefore reasonable to assume that the national study managers understood the importance of sampling and related factors prior to implementing the study.

To confirm that the IALS countries' sample designs met the stated criteria, each national study manager was required to submit a sampling plan to Statistics Canada. The plans were subjected to a thorough review by Mr. Georges Lemaitre, a private statistical consultant. All of the plans were judged to meet the sample design criteria, although numerous changes were negotiated with individual countries.

The balance of this chapter details the methodology used in each of the eight IALS countries in terms of target population and frame coverage and sample design.

2.2 Target Population and Frame Coverage

As noted earlier, each country participating in the IALS was required to draw a probability sample from which results representative of the civilian non-institutionalized population aged 16 to 65 could be derived. Countries were also permitted to sample older adults, and several did so. All IALS samples excluded full-time members of the military and people residing in institutions such as prisons, hospitals, and psychiatric facilities.

In six countries, the survey was carried out in the national language. In Canada, respondents were given a choice of English or French. In Switzerland, samples drawn from French-speaking and German-speaking cantons were required to respond in those respective languages. (Italian and Rhaeto-Romanic-speaking regions were excluded from this survey.) When respondents could not speak the designated language, attempts were made to complete the background questionnaire so that their literacy level could be estimated and the possibility of distorted results would be reduced.

Table 2.1 presents information on the ages sampled and language of the test for each IALS country. Table 2.2 reports the percentage of the population aged 16 to 65 covered in each country and lists excluded populations. As Table 2.2 makes clear, all of the IALS countries attained high levels of coverage. Thus, this aspect of the sample design would not contribute significantly to error in the final results.

Table 2.1: Ages sampled and language of test, by country

Country	Ages sampled	Language of test
Canada	16 and older	English French
France	16 to 65	French
Germany	16 to 65	German
Netherlands	16 to 74	Dutch
Poland	16 to 65	Polish
Sweden	16 and older	Swedish
Switzerland	16 and older	French German
United States	16 to 65	English

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 2.2: **Survey coverage, by country**

Country	Coverage (%)	Exclusions
Canada	98	residents of institutions, persons living on Indian reserves, members of the Armed Forces, residents of the Yukon and Northwest Territories, Francophone residents in the province of Ontario who lived in geographic regions where less than 20 persons were Francophone
France	95	residents of institutions, residents of mobile homes, residents of Corsica, residents living overseas
Germany	98	residents of institutions
Netherlands	99	residents of institutions
Poland	99	persons residing in Poland for less than 3 months
Sweden	98	residents of institutions (including those in military service), persons living abroad during the survey period
Switzerland	89	residents in Italian and Rhaeto-Romanish regions, residents of institutions, persons without telephones
United States	97	members of the Armed Forces on active duty, persons residing outside the U.S., persons with no fixed household address

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

More detailed information about the target populations, excluded groups, and sampling frames for individual IALS countries is provided below.

Canada

Canada's target population consisted of all household residents aged 16 and over. Excluded from the population were residents of the Yukon and Northwest Territories, residents of institutions, persons living on Indian reserves, and members of the armed forces. Also excluded were Francophone residents of the province of Ontario who lived in geographic regions where less than 20 persons were Francophone. Together these exclusions represent about 2 percent of the Canadian population aged 16 years and over. Canada's total in-scope population was 21,307,893 persons.

Canada used two separate frames to select its sample. The first was the 1991 Census file used to select the sample of Francophones from the province of Ontario. A separate literacy estimate for Franco-Ontarians was required, but because this population represented only about 5 percent of Ontario's population, a census frame was deemed most appropriate.

The second frame was the Labour Force Survey sample file used to select all other Canadians in the IALS sample. This file consisted of approximately 73,000 dwellings. After excluding dwellings found to be vacant, demolished or converted to non-residential uses, dwellings under construction, and seasonal dwellings, about 63,000 dwellings remained which were occupied by eligible persons. For these dwellings, the Labour Force Survey file contained demographic and labor force information for approximately 140,000 civilians aged 15 years and over.

France

France's target population consisted of all residents of ordinary households (using the Institute Nationale de la Statistique et des Etudes Economique, or INSEE, definition) aged 16 to 65 years inclusive. Excluded from this group were residents of Corsica, residents of institutions, and residents of mobile homes. These exclusions represented approximately 2.4 percent of the French population.

No exhaustive list of households exists for France. An area frame was therefore used.

Germany

In Germany, only those persons aged 16 to 65 years of age living in private households were included in the study. Residents of institutions, who represent less than 2 percent of the population, were excluded. Both East and West Germany were included, and no geographical regions were excluded. The total in-scope population was 53,826,289 persons.

An area frame was used to cover the target population.

Netherlands

The Netherlands target population consisted of those persons aged 16 to 74 years of age living in non-institutionalized dwellings. All residents of institutions (e.g. hospitals, nursing homes, homes for elderly, prisons, offices, businesses, government buildings, etc.) were excluded from the target population. These exclusions were estimated to be less than 1 percent. The total in-scope population of the Netherlands was 11,495,719 persons.

The initial sampling frame was a postal code file consisting of all postal codes in the country.

Poland

Poland's target population consisted of Polish citizens aged 16 to 65. Poland's total in-scope population was 24,475,650 persons.

Poland used the Polish National Register of Citizens (PESEL) as a frame for selecting the IALS sample. The PESEL covers all Polish residents living permanently (i.e., for longer than three months) in Poland. The register is continually updated for births, deaths, emigration, and change of permanent residence.

Sweden

The Swedish target population included all persons aged 16 years and over who were permanent residents of Sweden on 1 October 1994. Persons living in institutions, including those completing their military service, and persons living abroad during the survey period (1 October 1994 to 1 February 1995) were excluded. The total in-scope population of Sweden is approximately 6.9 million persons.

A register covering the total Swedish population, known as the "DAFA" register, was used as the frame.

BEST COPY AVAILABLE

Switzerland

The target population for Switzerland consisted of inhabitants of the country aged 16 years and over in the French- and German-speaking regions of the country. Excluded from this population were residents of the Rhaeto-Romanish and Italian speaking regions of the country (4.8 percent of the Swiss population) and residents of institutions (approximately 3.4 percent of the population).

An electronic telephone file was used as the sampling frame. This frame covers virtually the whole Swiss population since it is mandatory for persons with telephones to be registered on this file, and approximately 98 to 99 percent of all households have phones. The only exclusions from the frame are VIPs, whose numbers are not published for security reasons. A low percentage of households have more than one telephone line.

United States

The target population for the United States consisted of civilian non-institutionalized residents aged 16 to 65 years in the 50 states and the District of Columbia. Excluded from this group were members of the armed forces on active duty, those residing outside the United States, and those with no fixed household address (i.e., the homeless or residents of institutional group quarters such as prisons and hospitals). The total in-scope population of the United States was 165,301,676 people.

The sample for the American component of IALS was selected from persons in the Current Population Survey (CPS). The frame for the CPS consisted of 1990 Decennial Census files, which are continually updated for new residential construction and are adjusted for undercount, births, deaths, immigration, emigration, and changes in the armed forces.

2.3 Sample Design

As noted at the beginning of this chapter, each country participating in IALS was given explicit criteria to use in creating its sample design. These criteria specified that each country was to draw a high-quality probability sample representing the adult non-institutional civilian population aged 16 to 65. In the end, all of the countries secured samples that were large enough to support reliable proficiency estimates.

Table 2.3 summarizes the sample designs employed by the various IALS countries. As the table shows, all eight countries employed probability sampling for at least some stages of their sample designs, and five countries used probability sampling for all stages of sampling.

Three situations in which non-probability-based sampling methods were used are highlighted in the table. Specifically, France and Germany used "random route" procedures for selecting households into their samples, and Switzerland used an alphabetic sort to select one member of each household. Concerns that have been raised about these procedures are addressed later in this chapter.

BEST COPY AVAILABLE

Table 2.3: **Sample design, by country**

Country	1st stage of sampling	2nd stage of sampling	3rd stage of sampling	4th stage of sampling
Canada	Labour Force Survey, probability	LFS households, probability	1 person/household, probability	
France	communities, probability	blocks, probability	**households, non-probability**	1 person/household, probability
Germany	electoral districts, probability	**households, non-probability**	1 person/household, probability	
Netherlands	postal codes, probability	addresses, probability	1 person/household, probability	
Poland	cities/counties, probability	individuals, probability		
Sweden	municipalities, probability	individuals, probability		
Switzerland	telephone numbers, probability	**1 person/household, non-probability**		
United States	Current Population Survey, probability	CPS primary sampling units, probability	individuals, probability	

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The sample designs employed by the individual IALS countries are described below.

Canada

The Canadian Labour Force Survey (LFS) sample, from which the IALS sample was selected, was based upon a stratified, multi-stage design employing probability sampling at all stages of the design. The design principles were the same for each province. Each province consists of a number of economic regions, areas of similar economic structure formed on the basis of federal-provincial agreements. Each province is divided into Economic Regions, which are then organized into Self-Representing Units (SRUs) and Non-Self-Representing Units (NSRUs). SRUs are cities whose population generally exceeds 15,000 persons or whose unique characteristics demand their establishment as self-representing units. NSRUs are those areas lying outside the SRUs.

The use of LFS respondents was advantageous in that the LFS includes information about persons' ages and education, and this information made it possible to tailor the IALS sample to meet the survey requirements. The LFS population was further stratified using the age and education variables.

Literacy estimates were required at the national level as well as for six target groups. To provide reliable literacy estimates, a sample of at least 700 persons was allocated to each group. These target groups were:

i) In-school youth, aged 16 to 24

ii) Out-of-school youth, aged 16 to 24

BEST COPY AVAILABLE

iii) Seniors
iv) Unemployment Insurance Recipients
v) Social Assistance Recipients
vi) Francophones from the province of Ontario

In the Labour Force sample, SRUs and NSRUs are delineated into Primary Sampling Units (PSUs), which are areas that can be visited conveniently by an interviewer. A sample of PSUs is randomly selected. Selected PSUs are then delineated into clusters of dwellings which correspond to blocks or block faces. A sample of the clusters is selected, and all private dwellings in selected clusters are listed by field enumerators.

Within each selected cluster, a sample of dwellings is selected from the list of dwellings. Within each selected dwelling, labor force information is obtained for each civilian household member 15 years of age or older.

The IALS sample was selected as a subsample of this Labour Force Survey sample. A total sample of 6,427 persons aged 16 years and older was selected using stratified random sampling from the May 1994 LFS file. Only one person per household was randomly selected in order to keep the household response burden to a minimum.

The sample of Francophones from the province of Ontario was selected from 1991 Census files. The Census geographic areas were stratified according to whether they had a high, medium or low percentage of adult Francophones in their population, as of the 1991 Census. The sample was proportionately allocated into these strata based on the size of their Francophone populations.

The sample of Ontario Francophones was selected in a three-stage sample. In the first stage, equal-sized random samples of the geographic areas in Ontario were selected from the three strata (high, medium, and low density of Francophones). The second stage consisted of a systematic sample of 2,285 households from within the selected geographic areas. For each of the selected households, the interviewer determined which, if any, of the household members was Francophone and then selected one of these persons using the Kish random selection grid provided by Statistics Canada.

France

France's sample was selected using a three-stage sample design. Initially, the 21 regions of France were subdivided into areas according to five population size groupings: rural; areas with 2,000 to 20,000 inhabitants; areas with 20,000 to 100,000 inhabitants; areas with greater than 100,000 inhabitants; and the Paris region. In the first stage, communities from each region were randomly selected by population size cell, proportional to the size of the community. From the total 36,000 communities, 429 were selected into the sample.

Within each community, different sampling techniques were employed depending on the size of the community.

For communities with a population of more than 10,000 inhabitants, several steps were taken in order to obtain 24 addresses per community. First, specialized field staff randomly selected four different blocks with an assigned probability proportional to the number of phone lines in the block. Next, in each block, six addresses were selected using a random route method, then two sub-samples were created, each containing three addresses. These addresses were to be used by the interviewer at the next step of the sample selection process.

The telephone directory was used for the random route procedure. (In France, approximately 95 percent of households have a telephone. A small percentage of households have unlisted numbers or more than one number.) After a telephone listing was selected as the starting point, that household and the fifth and thirteenth households after it were selected into the sub-sample. To begin sampling for the second sub-sample, the interviewer was instructed to turn the corner onto the next street at the left and take one household within the third address on that street as the starting point, followed by the fifth and thirteenth households. Thus, at the end of this stage, the field staff had twelve pairs of addresses: two sub-samples of three from each of four city blocks within the sampling unit.

The field staff then selected seven pairs of addresses out of the initial set of twelve. The remaining ten addresses were to be used as replacements in order to meet the target of seven interviews per community.

The interviewer was instructed to first interview at the first address of each pair. If an interview was not possible, the second address was to be used as a replacement. If the second address again did not produce an interview, the interviewer approached the field staff to obtain a third (or fourth) replacement address.

For communities with populations of 10,000 or fewer inhabitants, the following steps were taken. The field staff randomly selected seven addresses from the telephone directory, then provided these addresses to the interviewers. Each interviewer was instructed to conduct the survey at the address following the one corresponding to the selected telephone number. If the selected address was in an apartment building, the interviewer would select the neighboring apartment with the first initial of the surname closest (alphabetically) to that of the originally selected address. If the selected address was an individual home, the interviewer would survey the next household. Specific guidelines were provided to the interviewers on how to select one replacement address, again using the random route method within communities of all sizes.

The interviewer verified that the selected household was not a vacant or seasonal dwelling. In a face-to-face interview, the interviewer drew up a roster of all household members and, for each person mentioned, collected data on name, birth date, sex, marital status, and relationship to the head of the household. This information was used to determine if at least one person aged 16 to 65 years lived in the household. If more than one individual in the house was in the target population, the interviewer selected one person to complete the IALS survey using the birthday method. Whichever person, aged 16 to 65, had a birthday closest to the date of the interview was selected.

Germany

In Germany, different master samples have been constructed through the cooperation of several market research agencies. Each master sample is representative of either East or West Germany and contains 210 electoral districts in each of which there are approximately 800 to 1,200 dwellings. Using systematic sampling, the electoral districts are selected for the master samples with probability proportional to the number of households. The electoral districts are classified by region, district, community size, district council, quarter, and vote area.

For IALS, two master samples were used in West Germany and a representative half-sample in East Germany. Interviewers were to complete five acceptable interviews per electoral district. To this end, interviewers selected 23 addresses beginning at a given starting point and following a predetermined random route. Because of the limited information available for each electoral district, the starting point was usually the first address, and the next 23 households were chosen for the initial sample. Then, nine of these addresses (numbered 1, 4, 6, 9, 11, 14, 16, 20, and 23) were selected for the sample. The final address was only to be used in difficult electoral districts. In actuality, it was used in only 18 of the 525 sampling units.

The interviewer verified that selected households were not vacant or seasonal dwellings. In a face-to-face interview, all eligible household members aged 16 to 65 were listed on a roster in order of decreasing age. In addition to recording the total number of eligible persons, the interviewers also listed their genders.

On each survey form, a random listing of the digits 1 - 9 (e.g., 437219856) was printed. The interviewer interviewed the person designated by the first digit in that series which was less than or equal to the number of eligible persons.

Netherlands

The initial sampling frame was a postal code file consisting of all postal codes in the country.

A sample of 3,000 successful interviews (2,750 persons aged 16 to 64 years, and 250 aged 65 to 74 years) was desired. An initial sample of 7,000 postal codes was selected to accommodate for estimated losses of 5 percent for nonexistent addresses, 10 percent for addresses outside of the target group, 15 percent for households in which the selected inhabitant was either outside of the target population or away for a long period of time and 40 percent for non-response.

A two-stage sampling approach was used to select the sample. In the first stage, a systematic sample of 7,000 postal codes was selected from a file of 540,817 codes. In the second stage, systematic sampling was again used to select one address per sampled postal code from the available list of addresses in each postal code. On average, there were 11.4 addresses per postal code. Large apartment buildings posed a problem, however, since they contained many addresses. A weight was therefore assigned to each address in the selected postal codes so that only one address per postal code was selected.

All selected addresses were mailed a letter informing the household that an interviewer would be visiting to conduct a literacy survey. A larger than expected number of letters were returned because they could not be delivered or because they had been sent to businesses. There had been no way to determine beforehand which addresses belonged to private households. Therefore, an additional sample of 2,000 addresses was drawn using the same techniques described above.

Within the selected households, the person in the target age range with the first-occurring birthday in the year was selected to participate in the survey. A counter was set up to count the number of elderly persons (65 to 74) selected. In theory, after conducting the predetermined number (250) of interviews for this age group, the remaining persons aged 65 to 74 years were to be excluded from the sample. In practice, this rule did not have to be implemented.

Poland

The Polish sample was based upon a stratified, two-stage design employing probability sampling at each stage of the design.

The population was first divided into 16 strata (eight urban and eight rural). The urban strata were constructed by grouping cities according to the number of inhabitants. The rural strata were constructed by collapsing counties, which are administrative units covering a number of villages or a city. Poland is divided into 3,009 counties, of which

856 are city counties and 2,153 are rural counties. The mean number of inhabitants in a rural county is about 6,800 persons.

The sample was distributed to each stratum with probability proportional to the number of inhabitants within the stratum. In the first stage of sampling, the sampling process was different for urban and rural strata.

There were two strata for urban areas with populations of 200,000 or more. One stratum contained five cities with populations of more than 500,000 (Warsaw, Lodz, Cracow, Poznan, and Wroclaw), and the second stratum contained 15 smaller cities with populations of 200,000 to 500,000. Each city was included in the sample for each strata.

For the remaining six urban strata, a cluster of 10 individuals per city was established. The number of cities was then selected from each stratum (randomly with replacement) so that the number of individuals selected (10 per city) would be equal to the sample size already allocated to that stratum. The ten individuals were selected systematically within each selected city.

A first-stage sampling strategy similar to the one used for the smaller urban strata was used for the rural strata. The primary sampling units for the rural strata were counties (administrative units covering a number of villages or a city). The cluster size used in rural strata was the same as that used in urban strata (10 individuals). The number of counties was then selected from each stratum (with probability proportional to size) so that the number of individuals selected (10 per county) would be equal to the sample size already allocated to that stratum. A systematic sample of individuals was then drawn independently in each selected county.

In urban and rural strata alike, individuals were sampled from the Polish National Register of Citizens (PESEL), a listing which covers all Polish residents living permanently (i.e., longer than three months) in Poland. Each individual listed in the PESEL has an assigned ID number which consists of the person's birthday (year/month/day) followed by five random digits. The PESEL is sorted by this ID number. A systematic sample was selected from this list. Because the file was already sorted by date of birth, the sample was automatically stratified by age.

In all, a total of 3,000 individuals (1,910 from urban strata and 1,090 from rural strata) were selected for the basic sample, and an additional 3,170 individuals (1,948 from urban strata and 1,123 from rural strata) were selected for a supplementary sample. The supplementary sample was selected to accommodate for non-response and was used for making substitutions in cases where non-response occurred. Both samples were stratified in exactly the same manner.

Sweden

Sweden's IALS sample was selected using a two-stage design. The 286 municipalities ("kommun") of the country were classified into two different categories: one defined as urban, consisting of all large urban municipalities, and the other defined as rural, comprising groupings of the remaining municipalities.

The urban municipalities were stratified into 30 strata depending on their population size. All 61 municipalities in the urban strata were selected for the sample. A random sample of persons from these municipalities was selected directly from the persons register using a systematic sample. The probability of selection, P, was given as:

P = total size of the sample / total size of the population

For the rural strata, one municipality was selected from each municipality grouping within a stratum. The probability of selection was proportional to the number of inhabitants who were age 16 years and older. The probability of selection of persons within the chosen municipalities was set so that the overall sampling probability, P, was exactly equal to the probability of selection of persons from the urban category. That is,

$$P_{person} = P / P_{municipality}$$

For each of the sampled municipalities, a random sample of persons was drawn from the Census register.

Switzerland

The target population was divided into two strata based on the official language (German speaking and French speaking) of each region. From the telephone frame, two systematic samples of addresses were drawn for each region: a main sample of 1,350 persons and a reserve sample of 500 persons. Each of these samples was divided into 2 sub-samples.

For the main sample, the first sub-sample was selected systematically within Cantons and District Councils. The second sub-sample was selected by taking the records that were 15 records down from the record selected in the first sub-sample. This method increased the probability of selecting at least two households in the same District Council.

The same method was used for the reserve sample.

Because separate estimates were required for the Canton of Geneva, a systematic sample of 600 telephone numbers in this area was drawn using the same method described above.

The final sample size, including the oversample for the Canton of Geneva, consisted of 3,000 persons.

The selected addresses were contacted by telephone. The person answering the call was asked to provide the number of household members who belonged to the 16-65 age group. The contact person was then asked for the given name of the age-eligible household member whose name came first in the alphabet, and this individual was selected for inclusion in the sample.

United States

The sample for the United States IALS was selected from a sample of individuals in housing units who were completing their final round of interviews in the Current Population Survey (CPS). The CPS is a large-scale continuous household survey of the civilian non-institutionalized population aged 15 and over. For the CPS, a sample of housing units is selected each month and is retained in the sample for four consecutive months. These units are then dropped from the sample for eight months, and then return to the sample for four additional months.

The United States IALS sample was selected from housing units undergoing their final interviews in March, April, May, and June 1994. These housing units were included in the CPS for their initial interviews in December 1992 and January, February, and March 1993. The IALS interviews were conducted in October and November 1994.

The CPS sample is selected using a stratified multi-stage design. Housing units that existed at the time of the 1980 Population Census were sampled from the Census list of addresses. Housing units that did not exist at that time were sampled from lists of new construction when available and otherwise by area sampling methods. Occupants of housing units that came into existence between the time of the CPS sample selection and the time of the IALS fieldwork had no chance of being selected for the IALS.

The IALS sample was confined to 60 of the 729 CPS primary sampling units (PSUs). Within these 60 PSUs, all persons aged 16 to 65 years of age in the sampled housing units were classified into 20 cells defined by race/ethnicity and education. Within each cell, persons were selected for the IALS with probability proportional to their CPS weights, with the aim of producing an equal probability sample of persons within cells. A total of 4,901 persons was selected for the IALS.

2.4 Concerns About Non-probability-based Sampling

The experts who reviewed the IALS methodology (see Appendix A) criticized the sample designs fielded in France, Germany, and Switzerland because not all of the final sampling units in these countries have known, non-zero probabilities of selection.

It is important to note, however, that the experts did not establish the fact that these procedures caused bias in the IALS estimates. They merely pointed out that bias could be present.

Bias caused by the Swiss selection method is probably low because it is unlikely that first name and literacy skill level are correlated. Questions concerning the random route methods applied at the final stage of sample selection in France and Germany are more complex. This section examines whether the procedures implemented in these countries are likely to have biased the estimates to a significant extent.

In particular, there has been concern that the interviewers' instructions for implementing the random route method were not detailed or precise enough to guarantee that all households in the primary sampling units had the same probability of being selected, as was assumed in weighting the survey data. In other words, the concern is that the sample weights assigned to sampled persons are not what they should be, in that some persons have a larger or smaller weight than is justified by their real probability of selection.

In a probability-based sample of persons from a particular population, every person in the population has a chance of being included in the sample, and the probability of inclusion is known. The actual sample represents one of many possible samples that could have been obtained from the population, using the particular sample design, and each such sample would generate an estimate x of a particular characteristic.

Although the term "unbiased estimate" is frequently used, unbiasedness is actually a property of an estimator or an estimation procedure, not an estimate per se. To say that an estimator is unbiased means that the average of the estimates produced from all possible samples equals the population value X. If we could actually produce estimates for all possible samples, we could calculate the standard deviation of these estimates about the population value. We know from sampling theory that approximately two-thirds of the estimates would be within one standard deviation of the population value, and approximately 95 percent would be within two standard deviations.[1]

---

[1] "Extreme" estimates are not just a theoretical possibility; they occur with some regularity in probability-based designs. We know that over repeated sampling, about 1 in 20 estimates of a particular characteristic will be at least two standard deviations away from the population value. The repeated sampling here could be the hypothetical repetition of the same sampling scheme in a single country, or the drawing of independent samples in different countries. In other words, if independent samples are drawn in 20 countries, then on average 1 country will have an estimate that is more than two standard deviations away from its population value and about 7 countries will have estimates that are more than one standard deviation away from their population values.

Next let us consider a sample drawn from the population, by whatever means, in which each person included in the sample is given the same weight, and the sum over all sampled persons equals the population total. In essence, this describes what was done in the IALS surveys in France and Germany, where random route methods were used in the primary sampling units.

In any sample design, samples can, in principle, be produced by a simple random sampling. If the weights of sampled persons are all the same, then the resulting estimates exactly equal those that would have been produced by corresponding simple random samples. Generally, we would view the estimates produced from simple random samples as good but would view exactly the same estimates produced by a flawed design as problematical. There is no contradiction here. The point is that bias is a meaningful concept only over repeated sampling, not for single estimates. The estimates from a flawed design may be "bad" on average, even if particular estimates from the design are quite good.

Let us now look more broadly at self-weighting designs, in which all sampled persons have the same weight. Such a sample design, even a flawed one in which the sample weights are incorrectly assigned as equal, may yield a set of samples that is a subset of all the possible samples that could have been drawn through simple random sampling. The question of interest is: What are the properties of the estimates produced from this subset of simple random samples? If the average of the estimates yields the population value, then the estimator based on the design is unbiased even if the sample weights are incorrect or certain persons in the population have no chance of being selected. In other words, even though there is no substitute for proper probability-based methods, non-probability based designs (and their resulting estimators) may not always have adverse consequences in practice.

The problem with non-probability-based designs is that we generally have no information about the properties of the resulting estimates, and therefore we cannot make any definitive statement about the quality of these estimates. In particular, we have no information concerning the probability that estimates will differ from the population value by specified amounts. The most we can say is that if the estimator is indeed biased, then the estimates will on average differ more from the population value than they would if the sampling weights had been properly assigned.

We now offer some practical considerations, with a view to determining if the sampling and estimation procedures in France and Germany are likely to have produced problematical estimates for the IALS.

For simplicity's sake, let us look at an extreme case: a flawed design in which certain persons in the population have no chance of being selected into the sample. If the characteristic to be measured is distributed in the same way among persons who can be selected as it is among persons in the population at large, then even the flawed design will yield unbiased estimates. In practice, of course, one has no way of knowing if this is the case. The question is, if a subset of the population has no chance of being sampled, how large and how different does that subset have to be in order to seriously distort the total population estimates, which are based on the portion of the population from which the sample is actually drawn? In this extreme case, the real probabilities of selection are unequal and unknown. In practice, all households in the sample might have a chance of being selected, but because of deficient sampling procedures (such as random route methods), the real probabilities of inclusion in the sample are not the same.

In Table 2.4, bold type denotes cases in which the population values for the effective sampled population and for the total population differ by about 5 or more percentage points.

Table 2.4: **Comparison of population and sample theoretic frequency**

	Relative frequency of characteristic (in %)		
Excluded population (percentage of total population)	Percent excluded in sampled population	Population value for the effective sampled population	Population value in total population
10	10	5	5.5
10	25	5	7.0
10	50	5	**9.5**
10	10	20	19.0
10	40	20	22.0
10	80	20	**26.0**
10	10	50	46.0
10	25	50	47.5
10	75	50	52.5
25	10	5	6.3
25	25	5	**10.0**
25	50	5	**16.3**
25	10	20	17.5
25	40	20	**25.0**
25	80	20	**35.0**
25	10	50	**40.0**
25	25	50	**43.8**
25	75	50	**56.3**

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 2.4 shows that a design must be seriously flawed in order for estimates based on the sampled population to differ substantially from the values for the total population. In a seriously flawed design, for example, a relatively large proportion of the population would be excluded from sampling, or the excluded population would be very different from the sampled population.

Examining the cases marked by bold type in Table 2.4, however, it is questionable whether the estimates are likely to produce a serious distortion of the true picture. For example, are we getting a significantly different picture of the literacy situation in a country if we are told, for example, that 20 percent of the population rather than 25 percent is operating at literacy levels 4 and 5? This discussion is not intended to excuse bad sampling. The point is that flawed sampling procedures do not necessarily change the picture that results from a proper sampling design. (A more stringent standard *would* be mandated, however, if one were measuring change over time; the accuracy of the measurement of change relies on accurate estimates, or constant biases, at multiple time points.)

Are distortions of the magnitude illustrated above likely in the IALS, given the nature of the sample designs implemented in the participating countries? The answer appears to be no. Although France and Germany used non-probability-based procedures for part of their sampling, the first stage of sampling in these countries involved the selection of 429 communes and 525 voter areas, respectively, using procedures that were entirely probability-based. The second stage of sampling in both countries involved selecting households based on random route methods. Since deficiencies in the random route procedures are associated with flawed implementation on the part of particular interviewers, the selection of a large number of primary sampling units introduces an element of randomization into the sample that is likely to attenuate the deficiencies in the random route procedures. In a number of the primary sampling units there may be large differences in either the real probabilities of selection or the population effectively sampled compared to the total population. It is, however, unlikely that this would occur systematically across many different sampling units drawn from all parts of a country, as would have to be the case to produce distortions of the magnitude shown earlier.

In Germany, the electoral districts in which households were selected by the random route method are relatively homogeneous, containing either houses or smaller apartment buildings, but not both. This reduces the possibility of bias.

Finally, both France and Germany benchmarked their IALS estimates for several dimensions (sex, age group, educational attainment, and occupation) to counts from the national census of the population for the same characteristics. This step, described at the end of Chapter 3, should largely correct for any sampling problems that would cause an effective sampled population to differ from the true population along one or more of these dimensions.

To sum up, although the sampling procedures used in some IALS countries were problematic, it is important to keep the impact of these problems on the quality of the IALS results in perspective. Clearly it is preferable to adhere to probability sampling methods. Nonetheless, for reasons outlined above, the deviations from strict probability sampling in France and Germany appear to be relatively minor and may not have seriously distorted the resulting survey estimates.

2.5 Conclusions

Each of the IALS countries attained a high level of population coverage. Thus, this aspect of the sample design would not contribute significantly to error in the final results.

In commenting on the survey coverage achieved by IALS, the experts' report indicates that the intended coverage of the IALS surveys conforms well to design specifications. What the report does not make clear is that, relative to other national and international comparative studies, the IALS coverage is extraordinary. Clearly, there is always room for improvement. To achieve substantial improvement, however, coverage deficiencies and other methodological problems must be addressed by the national statistics agencies upon which IALS countries rely for their data. It would be unrealistic to expect the IALS to be able to address problems that absorb much of the energies and attentions of these agencies.

The IALS countries were strongly encouraged to select probability samples representative of their population aged 16 to 65. The use of probability designs would make it possible to produce unbiased estimates for individual countries and to compare these estimates across the countries. Because the available data sources and resources were different in each of the participating countries, however, no single sampling methodology was imposed. Each IALS country created its own sample design.

Based on the available evidence, we do not believe that the sampling practices in Germany, France, and Switzerland, described earlier, introduced significant bias into the survey estimates. In Germany, where some units had no probability of selection, the possibility of bias depends on whether the geographic units that formed the final stage of selection were relatively homogeneous with respect to literacy. Comparisons of the distribution of the sampled population to external estimates over a range of characteristics suggest that the sample is indeed representative.

Nor is there a convincing case for significant bias in the French sample. Selection probabilities are known for the French sample but not explicitly to the final stage of selection, and once again, comparisons to external estimates over a range of characteristics suggest that the sample was representative. The French sample design employed a large number of small geographic units to drive the final stage of selection, a fact which strengthens the argument for homogeneity and therefore limits the potential for bias. The case for bias in the Swiss sample is the most difficult to evaluate, although the Swiss demographic statistics also line up with estimates from independent sources.

The preceding review makes it clear that the quality assurance procedures for sample design in the IALS were inadequate in that three countries used less than ideal design parameters. Three improvements are implied for future rounds. First, Statistics Canada must provide more detailed design specifications to guide future data collection efforts. This improvement has already been accomplished. Second, a much more extensive review and documentation of planned sampled designs must be undertaken in advance of fielding the main study. Once again, this step has already been implemented in the current round of data collection. Finally, Statistics Canada, in its role as study manager, must retain the authority to qualify or to refuse to publish comparisons of national estimates where the quality of the underlying sample is suspect.

Chapter 3:

Survey Response and Weighting

Nancy Darcovich

3.1 Introduction

One of the greatest challenges that faces a large-scale survey such as the IALS is to achieve a high response rate—to collect information from as many sampled individuals as possible. Typically, some individuals who are selected to respond to a survey cannot be contacted, and among those who are contacted, some refuse to participate and others cannot provide the information requested. In the IALS, securing a high rate of response was expected to be challenging for all of the participating countries because the survey interview was lengthy and because sampled individuals were asked to take a test.

A survey's sample size can be inflated at the design stage to compensate for an anticipated rate of non-response, but this procedure does not reduce bias resulting from non-response. Non-response bias occurs when survey respondents and non-respondents differ in important respects. The magnitude of the bias is a function of the size of the non-response and the differences in characteristics between respondents and non-respondents.

This chapter provides information that can be used to explore the possibility of non-response bias in the IALS survey. The first part of the chapter presents information about survey response in the participating countries and describes steps taken by each country to obtain a high rate of response. The latter part of the chapter describes the different methods that the countries used to weight their data. Comparisons of the sampled population with the weighted population estimates are also provided.

3.2 Survey Response

The IALS instruments consisted of three parts:

i) a background questionnaire, which collected demographic information about respondents;

ii) a set of core literacy tasks, which screened out respondents with very limited literacy skills; and

iii) a main booklet of literacy tasks, which was used to calibrate literacy levels.

The definition of a respondent for the IALS is a person who partially or fully completed the background questionnaire. In some cases, incomplete assessment data were obtained, but if the individual provided background information and indicated why he or she did not complete the core and main literacy task booklets, it was possible to impute a literacy profile for the person (see Chapter 11 for more detailed information).

As explained in Chapter 2, each IALS country was required to provide a sample of 1,000 survey respondents. This sample size ensured sufficient information for the calculation of reliable literacy profiles. To permit international comparisons, each country was required to cover ages 16 to 65 inclusive. The numbers of respondents for the various countries are presented in Table 3.1.

All countries participating in the IALS instructed interviewers to make call-backs at households that were difficult to contact. Given the length of the IALS interview, however, as well as the fact that part of the interview involved taking a test, one might expect response rates to be low. As shown in Table 3.2, response rates ranged from a high of 75 percent (Poland) to 45 percent (the Netherlands). This table also indicates the ages sampled and the total number of respondents (all ages) for each country.

Table 3.1: **Survey coverage, language of test, and number of respondents, by country**

Country	Population covered*	Language of test	Number of respondents*
Canada	13,676,612 4,773,648	English French	3,130 1,370
France	36,432,000	French	2,996
Germany	53,826,289	German	2,062
Netherlands	10,460,359	Dutch	2,837
Poland	24,475,649	Polish	3,000
Sweden	5,362,942	Swedish	2,645
Switzerland	1,008,275 3,144,912	French German	1,435 1,393
United States	161,121,972	English	3,053

* Aged 16 to 65.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 3.2: **Ages sampled, number of respondents, and response rate by country**

Country	Ages sampled	Number of respondents	Response rate (%)
Canada	16 and older	5,660	69
France	16 to 65	2,996	55
Germany	16 to 65	2,062	69
Netherlands	16 to 74	3,090	45
Poland	16 to 65	3,000	75*
Sweden	16 and older	3,038	60
Switzerland	16 and older	3,000	55
United States	16 to 65	3,053	60

* Poland's response rate includes only the first wave of sampled persons.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Response information for the various IALS countries is provided below.

Canada

When sampled persons were hard to reach, Canada's interviewers were instructed to call back at least three times, always at different times of the day or evening.

Because the main Canadian sample was selected from respondents to the Labour Force Survey, the total non-response rate for IALS includes non-response to the LFS as well as non-response to IALS. The response rate for the LFS in May 1994 was 92.1 percent. The largest group of non-respondents to the LFS is males between the ages of 20 and 24 inclusive (1.4 percent of the non-response). However, LFS estimates are weighted up to population projections from the 1991 Census by province, sex, and age group cells to adjust for undercoverage.

Of the 6,427 persons selected in the main IALS sample, 4,703 (73.2 percent) responded to the IALS survey. Thus, the overall response rate was 67.4 percent when the LFS response rate was included.

For the Franco-Ontarian portion of the survey, the initial sample consisted of 2,285 people in the same number of households. The first stage of data collection involved contacting any member of the selected household to determine whether the household was in scope. There was a successful contact in 85.6 percent (1,956) of the cases. Non-response at this stage was divided equally between individuals who refused outright to participate in the survey and those whom interviewers were unable to contact during the entire data collection period. The second stage of data collection involved determining household eligibility. Nearly 40 of the households sampled proved to be ineligible because there were no Francophones living there. The response rate among persons selected from eligible households was 88.2 percent. Together, the household (85.6 percent) and individual rates (88.2 percent) yield an overall response rate of 75.5 percent for the Franco-Ontario portion of the Canadian IALS. Combining the two samples meant that records that belonged to Franco-Ontarians from the LFS frame needed to be dropped. Thus, the final sample was 5,660 respondents aged 16 years and older (4,616 from the LFS frame and 1,044 from the Census frame) with a final combined response rate of 69 percent.

France

France's desired sample size was 3,003 persons. The data collection contractors actually achieved 99.8 percent of this goal by using replacement dwellings when non-response was encountered at originally selected households.

The following table shows the number of times a replacement was used instead of the initial address selected.

Address	Number	Percent
Initial	1,363	45.5
1st replacement	792	26.4
2nd replacement	841	28.1
Total	**2,996**	**100.0**

It should be noted that the French interviewers followed their procedures closely and did not use the first replacement until a hard refusal was obtained from the initial address. Four attempts had to be made at different times of the day, or an outright refusal given, before a refusal could be coded.

Germany

Germany's interviewers were instructed to attempt to contact a selected household at least three times at different times on different days before they coded a household as a non-response.

Germany obtained a response rate of 69 percent. There was, however, a very small percentage of households in which language difficulties were cited. This percentage seems to agree with the results of another recent survey, "ZUM - ALLBUS."

Netherlands

In the Netherlands, interviewers were instructed to call back at least four times before coding a household as a non-response.

The Netherlands' total initial sample size was 9,000 households. Interviews were actually completed for 3,090 households, while 3,811 were non-respondents and 2,099 were out of scope. Of the 2,099 out-of-scope records, 809 were so identified after the first wave of introductory letters was mailed, 460 were identified as businesses or institutions by the interviewers, and 830 contained respondents who were outside the target age group. The 830 out-of-scope respondents represent 9.2 percent of the initial sample.

With an in-scope sample of 6,901 households (9,000 minus 2,099), one interview per household, and 3,090 completed interviews, the Netherlands had a response rate of 44.8 percent. It should be noted that Dutch surveys generally have high rates of non-response, perhaps because a large number of interviews are conducted in the country on an annual basis. Experience shows that most Dutch social surveys encounter serious response problems.

Poland

In attempting to contact hard-to-reach households, Poland's interviewers were instructed to call back at least three times and attempt to schedule an interview time that was convenient for the respondent.

Poland's initial sample size was 6,071 persons. This number included an initial sample of 3,000 and a second sample of 3,071 replacements to be used if the originally selected person did not respond. The actual number of survey respondents was 3,000.

Poland reports that 24.1 percent of the originally sampled households were non-responsive, so replacement households needed to be contacted. No report of the number of replacement households that subsequently refused was given, however. Thus, the response rate of 75.9 percent represents only the rate after the first wave of interviewing.

Sweden

The Swedish interviewers were instructed to make a minimum of 10 calls and site visits if they experienced a problem contacting a selected respondent. Respondents who were immigrants with a limited command of Swedish were informed that the background questionnaire could be answered with the assistance of a family interpreter.

Sweden's initial sample size was 5,275 persons, 211 of whom were out of scope. Thus, the net sample deemed to be in scope was 5,064. While 3,038 (60.0 percent) answered the background questionnaire, there were 2,026 non-respondents.

Switzerland

Switzerland selected an initial sample of 3,000 persons as well as a second sample to replace non-responding cases.

In the German part of the sample, a total of 3,043 interviews were attempted. Because 694 addresses were out of scope (e.g., business address, fax number, summer home), the total in-scope sample was 2,349 addresses. In all, 1,399 successful interviews were conducted, resulting in a response rate of 59.6 percent.

Similarly, in the French portion of the sample, a total of 3,728 interviews were attempted. Because 856 of these addresses were out of scope, the total in-scope sample was 2872 addresses. Successful interviews were completed with 1,444 individuals, for a response rate of 50.3 percent.

Averaging the two response rates of 59.6 percent and 50.3 percent, Switzerland reported an overall response rate of 55.0 percent.

United States

In the United States, non-response to the IALS occurred for two reasons. First, some individuals did not respond to the Current Population Survey (CPS). Second, some of the CPS respondents selected for IALS did not respond to the IALS instruments. In any given month, non-response to the CPS is typically quite low, around 4 to 5 percent. Its magnitude in the expiring rotation groups employed for IALS selection is not known. About half of the CPS non-response is caused by refusals to participate, while the remainder is caused by temporary absences, other failures to contact, inability of persons contacted to respond, and unavailability for other reasons.

A sizable proportion of the non-response to the IALS background questionnaire was attributable to persons who had moved. For budgetary reasons, it was decided that persons who were not living at the CPS addresses at the time of the IALS interviews would not be contacted. This decision had a notable effect on the sample of students, who are sampled in dormitories and other housing units in the CPS only if they do not officially reside at their parents' homes. Those who reside at their parents' home are included in the CPS at that address, but because most of these students were away at college during the IALS interview period (October to November 1994), they could not respond to the IALS.

The high level of non-response for college students could cause a downward bias in the literacy skill-level estimates. This group represents only a small proportion of the United States population, however, so the potential bias is likely to be quite small. Further, comparison of the IALS results to the U.S. National Adult Literacy Survey data discounts this as a major source of bias.

Out of the 4,901 CPS respondents sampled for IALS, 3,060 responded. Including a non-response rate of approximately 5 percent from the CPS, a total response rate of 59.4 percent (62.4 x 95.0) was obtained for the United States IALS.

3.3 Weighting

IALS countries used different methods for weighting their samples. For countries with known probabilities of selection, a base weight calculated using the probability of selection can be computed. Another method commonly used to weight data, discussed later in this chapter, is to adjust the rough estimates produced by the sample to match known population counts from non-IALS sources.

The weighting procedures used by each individual country are discussed below.

Canada

LFS weighting

Since the main Canadian IALS used a subsample of the LFS sample, the derivation of weights for the survey records is directly tied to the weighting procedure used for the LFS.

In the LFS, the final weight attached to each record is the product of a basic weight, a cluster sub-weight, a balancing factor for non-response, a rural-urban factor, and a province-age-sex ratio adjustment factor. Each is described below.

Basic weight

In a probability sample, the sample design determines weights that must be used to produce unbiased estimates for the population. Each record must be weighted by the inverse of the probability of selecting the person to whom the record refers. Because all eligible individuals in a dwelling are interviewed (directly or by proxy), this probability is essentially the same as the probability with which the dwelling is selected.

Cluster sub-weight

The cluster delineation is such that the number of dwellings in the sample increases very slightly with moderate growth in the housing stock. In clusters where substantial growth has occurred, sub-sampling is used to keep interviewer assignments manageable. The cluster sub-weight is the inverse of this sub-sampling ratio in clusters where sub-sampling has occurred.

Non-response

Notwithstanding the strict controls of the LFS and the interviewers' attempts to contact sampled persons, some non-response is inevitable. The LFS non-response rate is approximately 5 percent. For certain types of non-response (e.g., household temporarily absent, refusal), data from a previous month's interview with the household, if available, are brought forward and used as the current month's data for the household. In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is made separately for geographic areas called balancing units. It assumes that the households that are interviewed represent the characteristics of those that should have been interviewed. The estimates will be somewhat biased to the extent that this assumption is not true.

Rural-urban factor

In geographic areas without sufficient rural and urban populations for explicit urban and rural strata to be formed, each primary sampling unit (PSU) is composed of both urban and rural parts. Information about the total population in rural and urban areas is available from the 1991 Census for each PSU as well as for each economic region (ER) in which explicit urban/rural stratification is not performed. Comparison by ER with the actual 1991 rural or urban census counts indicates whether the selected PSUs over- or under-represent the respective areas. The ratio of actual rural-urban counts is divided by the corresponding estimates. These two factors are computed for each relevant ER at the time the PSUs are selected and are entered on each sample record according to the appropriate description (rural or urban) of the geographic area. Changes in these factors are incorporated at the time of PSU rotations.

Subprovincial and province-age-sex adjustments

By applying the previously described four weighting factors, an initial weight or "subweight" is calculated. Valid estimates can be derived for any characteristic for which information is collected by the LFS—in particular, estimates of the total number of persons age 15 and older in provincial economic regions and the 24 large metropolitan areas, and estimates of designated age-sex groups in each of the 10 provinces.

Independent estimates are available monthly for each of these classes from projections based on the 1991 Census counts. By using an interactive "raking ratio" adjustment procedure, the weights derived to this point are adjusted by a multiplying factor to correspond to the independent estimate for the various classes. This factor is the ratio of the independent estimate to the survey estimate based on the first four weighting factors. This final adjustment ensures that basic provincial and total population counts for economic regions, along with age/sex distribution data published from the LFS, correspond to other Statistics Canada data sources.

The adjustment also increases the precision of all estimates derived from the LFS. This final adjustment is known as the "final weight" of the LFS.

IALS weighting

The principles used in calculating weights for the Canadian IALS main sample are identical to those for the LFS. Three adjustments were made to the LFS weights in order to derive final weights for the individual records on the file:

(1) An adjustment to reflect the selection probabilities associated with the IALS sub-sample,

(2) An adjustment to account for the additional non-response to the main Canadian IALS, and

(3) Re-adjustments to account for independent province/age/sex projections and for independent Economic Region/Census Metropolitan Area projections, made after the above adjustments.

In the Franco-Ontario sample, the final weight of every record is the product of five numbers: the weight for the first stage of selection, an adjustment factor for this stage, the weight for the second stage of selection, an adjustment factor for this stage, and the weight for the third stage of selection. These are defined below.

The weight for the first stage of selection is the inverse of the probability of selection of the primary sampling unit (PSU). It varies according to the size of the PSU.

The first adjustment factor corrects for complete non-response at the PSU level. The adjustment factor is the ratio of total PSUs selected to responding PSUs and is done independently in strata 1 and 2. In practice, there was no complete non-response, and all the factors ended up being equal to unity.

The weight for the second stage of selection is the inverse of the probability of selection of the household. All households in a given PSU have the same weight (provided they belong to the target population of households).

The second adjustment factor has two purposes: to correct for non-response for households that belong to the target population, and to correct for out-of-scope households that do not. The adjustment factor is simply the ratio of total households selected to responding households, and is calculated at the PSU level. Correcting for out-of-scope households means accounting for households that did not include Francophones at the last census but into which Francophones have moved since. The adjustment ensures that the number of households with Francophones is comparable to what it was at the last Census.

The weight for the third stage of selection is the inverse of the probability of selection of the individual. In other words, the weight is equal to the number of Francophones aged 16 or over in the household.

The next step in the weighting procedure was to merge the data from the LFS-based and census-based samples and to adjust the weights accordingly.

First, Franco-Ontarians in the LFS-based sample were identified and eliminated from the sample. These already-weighted records were then replaced by the 1,440 respondents from the census-based sample. The resulting file size was 5,660 records. Next, the weights were adjusted to comply with the French/English speaking counts for Ontario. Finally, these weights were adjusted based on the province/sex/age categories used previously. The resulting final weight appears on the IALS microdata file.

France

The French data were weighted based on a comparison of the IALS sample-based estimates to other known population counts. Six characteristics were taken into account, in sequence: total number of persons in the household, number of household members between 16 and 65 years of age, education (using March 1995 Labour Force Survey counts), size of community, age, sex, and occupation. An iterative program was used to reduce the gap between the estimates and the external sources.

Germany

The German data were weighted by comparing the estimates taken from the sample with external sources. The sample was split into 277 different sub-groups defined by the intersection of different characteristics, including the number of household members aged 16 to 65, age and sex, and citizenship.

Netherlands

Each respondent in the Netherlands sample was assigned a "base weight." This weight was calculated by dividing the in-scope target population size by the number of respondents. The Netherlands team then adjusted its sample counts to correspond to 1994 counts from the Central Bureau of Statistics (CBS) for Dutch inhabitants aged 16 to 74. These adjustments were based on four demographic characteristics: region, age, sex, and education.

Poland

The Polish team used age groups, rural-urban codes, and regions (voivodeships) as post-stratification variables. This information was derived from Poland's Statistical Yearbook of Demography (Central Statistical Office, Warsaw, 1993). Although this benchmarking information was slightly out-of-date, Poland's demographic counts are known to be stable over a two- to three-year period.

Sweden

Sweden employed a self-weighting sample design. Estimates were benchmarked to geographic region, educational attainment, age, and sex counts.

Switzerland

Switzerland used a number of benchmark characteristics to adjust the sample data, with one reweighting characteristic after another used to adjust the sample. After each reweighting, certain characteristics were checked to ensure that the results were not biased by the procedures. The six characteristics used by the Swiss were: number of household members aged 16 to 65, total number of persons in the household, level of education, size of community, age, and sex.

United States

Weights for the United States IALS included two components. The first assigned weights to Current Population Survey (CPS) respondents, and the second assigned weights to the IALS respondents.

The CPS weighting scheme is a complex one involving three components: basic weighting, non-interview adjustment, and ratio adjustment. The basic weighting compensates for unequal selection probabilities. The non-interview adjustment compensates for non-response within weighting cells created by clusters of PSUs of similar size; Metropolitan Statistical Area (MSA) clusters are subdivided into central city areas, and the balance of the MSA and non-MSA clusters are divided into urban and rural areas. The ratio adjustment makes the weighted sample distributions conform to known distributions on such characteristics as age, race, Spanish origin, sex, and residence.

The IALS sample design began with persons aged 16 to 65 in the 60 sampled PSUs and the persons' final CPS weights. Individuals were sampled for IALS with probabilities proportional to their final CPS weights. The weights of persons sampled for IALS were adjusted to compensate for the use of the four rotation groups, the sampling of the 60 PSUs, and the sampling of persons within the 60 PSUs.

The IALS non-interview adjustment compensated for sampled persons for whom no information was obtained because they were absent, refused to participate, had a short-term illness, had moved or had experienced an unusual circumstance that prevented them from being interviewed.

Finally, the IALS ratio adjustment ensured that the weighted sample distributions across a number of education groups conformed to March 1994 CPS estimates of these numbers.

3.4 Benchmarking

One commonly used approach is to weight survey data is to adjust the rough estimates produced by the sample to match known population counts from sources external to IALS. This "benchmarking" procedure assumes that the characteristics of non-respondents are similar to those of respondents. It is most effective when the variables used for benchmarking are strongly correlated with the characteristic of interest—in this case, literacy levels.

For the IALS, the key benchmarking variables were age, employment status, and education. As shown in Table 3.3, all of the IALS countries benchmarked to at least one of these variables. Five used education, the variable most strongly correlated with literacy.

Table 3.3: **Benchmark variables by country**

Country	Benchmark variables
Canada	Province, economic region, census metropolitan area, age, sex, in-school youth, out-of-school youth, unemployment insurance recipients, social assistance recipients
France	total number of persons in the household, number of household members between 16 and 65, education, size of community, age, sex, profession
Germany	number of household members aged 16 - 65, age, sex, citizenship
Netherlands	region, age, sex, education
Poland	region, urban-rural, age
Sweden	region, education, age, sex
Switzerland	number of household members aged 16 - 65, total number of persons in the household, level of education, size of community, age, sex
United States	education

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994

3.5 Internal and External Checks

To determine how similar or different the sampled persons were from the overall population, researchers compared the distribution of the sample and the weighted sample by age and sex (Table 3.4). These two variables were selected because most countries post-stratified their samples to produce reliable estimates for these two variables.

Table 3.4: **Comparison of sample and weighted distributions of age/sex groups, by country**

Age /Sex group distribution	Canada	France	Germany	Netherlands	Poland	Sweden	Switzerland	United States
Age 16 to 24								
Male								
Sample	12	8	7	6	11	10	6	9
Weighted	9	10	8	9	11	10	9	8
Female								
Sample	14	9	7	6	10	9	6	9
Weighted	9	10	7	9	10	10	6	8
Age 25 to 44								
Male								
Sample	19	21	21	26	21	21	25	21
Weighted	24	22	22	24	24	21	24	21
Female								
Sample	25	28	28	29	24	22	28	24
Weighted	25	23	24	23	24	21	24	25
Age 45 to 65								
Male								
Sample	13	15	17	17	16	18	17	17
Weighted	16	17	19	17	15	18	17	18
Female								
Sample	16	18	20	17	8	20	19	20
Weighted	17	17	20	17	16	19	20	20

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994

The table shows that although France, Germany, and Switzerland selected non-probability samples at one stage in their sampling process, the age and sex characteristics of their samples are similar to the final weighted distributions post-stratified to age and sex.

Differences do appear in the distributions for Canada, but Canada purposely oversampled its younger population in order to produce reliable estimates for the 16 to 24 age group. The differences in sampling probabilities were taken into account in Canada's weighting process.

Younger males and 25 to 44 year old females were under-represented in the Dutch sample. Because researchers in the Netherlands post-stratified to age, sex, and education counts, however, they would have adjusted for this bias in the weighting process.

Because level of education correlates strongly with literacy, the extent to which the IALS sample estimates for education level compare with other information about education level provides another indication of how well the IALS sample reflects its population. Table 3.5 compares the educational attainment of the age 25 to 64 population as reported in OECD's *Education at a Glance* (1992) to the educational attainment estimates obtained from the IALS samples using the final weights.

Most of the IALS estimates of educational attainment closely match those of the OECD. The differences in educational coding and reference year between the two surveys, combined with the sampling error from both surveys, is probably enough to account for the differences. Thus, these data do not provide significant evidence of any bias in the IALS results due to non-response.

Table 3.5: Comparison of educational attainment estimates from OECD and IALS for persons aged 25 to 64

Country	Estimate	Below upper secondary	Upper secondary	Non-university tertiary	University
Canada	OECD	29	30	26	15
	IALS	32	32	18	18
France	OECD	48	36	6	10
	IALS	42	40	7	11
Netherlands	OECD	42	37	*	21
	IALS	43	36	0	21
Poland	OECD	not available			
	IALS	61	24	7	9
Sweden	OECD	30	46	12	12
	IALS	27	44	15	14
Switzerland	OECD	19	60	13	8
	IALS	20	60	11	9
United States	OECD	16	53	7	24
	IALS	14	46	15	25

* Netherlands does not use ISCED 05 in reports on its education system.

Sources: IALS data: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994. OECD data: OECD, *Education at a Glance* 1992.

3.6 Sampling Error

As well as supplying the weights for their sample, each country was also responsible for providing the information necessary to run variances for their literacy proficiency estimates. Each country was to provide a set of replicate weights for use in a Jackknife variance estimation procedure. This is a procedure which is able to handle both complex sampling designs and complex estimators. It works by calculating variance estimates using subsets of the sample and then combining these estimates to form an overall variance estimate.

3.7 Conclusions

Given the length of the survey interview and the fact that it involved taking a test, obtaining a high response rate was expected to be a challenge for the IALS countries. In fact, response rates varied considerably from country to country, ranging from a high of 75 percent (Poland) to 45 percent (the Netherlands). In an effort to maximize the response rates, all of the IALS countries instructed interviewers to make call-backs at households that were difficult to contact.

The countries used different methods to weight their samples. All post-stratified their data to known population counts, and five post-stratified to education variables, which are strongly linked to literacy. Post-stratification compensates for non-response, assuming that respondents and non-respondents are the same conditional on the characteristics for which they are being adjusted.

A comparison of the distribution of the age and sex characteristics of the actual and weighted samples indicates that the samples were comparable to the overall populations of the IALS countries. Although France, Germany, and Switzerland selected non-probability samples at one stage in their sampling processes, the distributions of their samples across age and sex groups is similar to the final weighted distributions. Further, although younger males and 25- to 44-year-old females were under-represented in the Netherlands sample, post-stratification to age, sex, and education counts adjusted for this bias in the weighting process. Differences in population distributions in the data for Canada resulted from an intentional oversampling of younger adults and were taken into account in the weighting.

In closing, while the experts' report described the non-response rates in IALS as disturbing and noted that the results may therefore be subject to serious bias, two points must be emphasized.

First, expectations for response should be reasonable. IALS was a demanding survey. On average, respondents were asked to donate 30 minutes of their time to answer background questions and another hour to take a test, a task with which few would have had recent experience. The IALS response rates are better than those achieved by many other international comparative studies as well as by many national labor force surveys in Europe whose data are widely published and used. It would therefore be unfortunate to impose unrealistic expectations concerning response rates on future rounds of the study.

Second, there is little actual evidence of non-response bias in the IALS results. Earlier Canadian and American studies show that fully 85 percent of the observed variance in literacy skill profiles can be accounted for by a handful of variables, including parents' education and respondent's education, age, and mother tongue. Even a cursory analysis of the variation in these determinants across IALS countries (see Chapter 12) suggests that the literacy profiles are highly plausible. Further, the statistical analyses presented in Chapters 10 and 11 provide little, if any, evidence of bias between or within countries.

Item response theory (IRT) and the imputation method are highly sensitive instruments for detecting bias associated with differential item functioning. To go undetected, a bias would have to be in the same direction and magnitude across the entire spectrum of ability in a country, and the same phenomenon could not be operating in other countries. The psychometric evidence for all the IALS countries suggests that tested proficiency behaves in a predictable and stable way across the entire demographic spectrum.

Chapter 4:

Non-response Bias

Nancy Darcovich, Marilyn Binkley, Jon Cohen,
Mats Myrberg, and Stefan Persson

4.1 Introduction

As noted in the previous chapter, a post-hoc evaluation of the IALS methodology raised the possibility of non-response bias in the study results. Three studies were subsequently conducted to investigate this matter. Because the sampling frames for Canada and the United States contained information about the characteristics of sampled individuals, it was possible to compare the characteristics of respondents and non-respondents, particularly with respect to literacy skill profiles. Sections 4.2 and 4.3 summarize the results of the Canadian and United States investigations, respectively. Section 4.4 present the results of a non-response follow-up study commissioned by the Swedish National Study Team. Collectively, these studies shed additional light on the possible magnitude of non-response bias present in the IALS estimates.

4.2 Canadian Study of Non-response Bias

The investigation of non-response bias in the Canadian IALS proceeded in three stages:

- Stage 1: Examine the demographic characteristics of respondents and non-respondents to compare response probabilities across demographic groups;
- Stage 2: Compare estimates from the sample using the base weights with estimates derived using the final weights after post-stratification and adjustment for non-response; and,

- Stage 3: Determine an acceptable level of potential bias, then calculate the size of the average difference between respondents and non-respondents that would be required to cause bias of this magnitude.

The Canadian IALS consisted of a main sample as well as a special sample of Francophones living in the province of Ontario. The non-response study was conducted using only the main sample, because demographic information was available only for this sample. The main sample represents 75 percent of the total Canadian sample selected and 98 percent of Canada's target population.

The main sample for the Canadian IALS was selected from a file of persons aged 16 and older who responded to the May 1994 Labour Force Survey (LFS). Using demographic information contained in this file, it is possible to compare the characteristics of IALS non-respondents with those of respondents.

The LFS had a response rate of 92.1 percent for May 1994, the month from which the sample was drawn. From the main sample of 6,427 persons selected to participate in IALS, a response rate of 73.2 percent was obtained. Thus, overall, a 67.4 percent response rate was achieved. The overall response rate for the section of the Canadian IALS sample studied was 72.9 percent (discounting non-response to the LFS). From an initial sample of 5,081 persons, there were 3,706 respondents.

Studies have shown that the largest group of non-respondents to the LFS in May 1994 was males between the ages of 16 and 24 (18.2 percent of the non-respondents). The second largest group was males in the 25 to 44 age group (14.9 percent). This underrepresentation was accounted for by benchmarking (post-stratification) to age-sex projections from the 1991 Canadian Census of Population.

Researchers studied the age, education, and employment characteristics of individuals who responded to the LFS but who did not respond to the IALS. These variables were chosen because the IALS results indicated that they are highly correlated with literacy skills. The analysis was restricted to persons aged 16 to 65, since this age range was covered by all IALS countries. The age groups used for the analyses correspond to those used in IALS publications.

Stage 1

Table 4.1 presents response rates for the Canadian main sample by age group. Table 4.2 presents response rates by employment status and education level. No meaningful differences in response rate are found among adults in various age groups or among groups defined by employment status and level of education.

Table 4.1: **Response rates by age group**

Age	Sample size	Number of respondents	Response rate (%)
16 to 35	2,771	2,003	72.3
36 to 55	1,693	1,259	74.5
56 to 65	617	444	72.0
Total	**5,081**	**3,706**	**72.9**

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 4.2: **Response rates by employment status and education**

Employment status	0 - 8 years	Some secondary	Completed secondary	Some post-secondary	Completed post-secondary	Total
Employed						
Sample	196	589	596	295	1,198	2,874
Response rate (%)	73.5	70.6	75.5	71.2	73.1	72.6
Unemployed						
Sample	65	158	87	58	176	544
Response rate (%)	72.3	70.0	73.6	75.9	68.2	70.6
Other						
Sample	353	632	248	126	304	1,663
Response rate (%)	77.9	72.5	74.6	67.5	76.3	74.5
Total						
Sample	614	1,379	931	479	1,678	5,081
Response rate (%)	75.9	71.3	75.1	70.8	72.6	72.9

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Next, researchers examined response rates for groups defined by the three-way interaction among age, employment status, and education characteristics. The goal was to see whether any of the cell response rates differed significantly from the overall response rate.

Table 4.3 presents these data for adults in three age groups: 16 to 35, 36 to 55, and 56 to 65. Cells in which the response rate was much lower (at least 8 percent lower) than the overall response rate appear in bold type. A drop of 8 percent was defined as representing a meaningful change. Response rates based on an initial sample size of fewer than 20 persons are marked with an asterisk to indicate that the data are not reliable.

Table 4.3: **Response rates by age group, employment status, and education**

Employment status	0 - 8 years	Some secondary	Completed secondary	Some post-secondary	Completed post-secondary	Total
Age 16 to 35						
Employed	70.4	73.6	74.7	69.7	71.4	72.4
Unemployed	80.0	**64.8**	77.2	74.4	**63.0**	68.8
Other	73.2	75.2	69.6	73.0	71.9	73.5
Total	73.3	73.3	73.9	71.1	70.4	72.3
Age 36 to 55						
Employed	74.7	68.5	77.8	76.5	74.0	74.1
Unemployed	72.7	78.0	66.7*	85.7*	73.5	74.3
Other	76.8	**65.0**	82.5	62.5*	80.0	74.9
Total	75.5	68.6	78.0	74.5	74.8	74.5
Age 56 to 65						
Employed	74.5	**51.2**	66.7*	66.7*	68.1	65.8
Unemployed	42.9*	77.8*	66.7*	0*	87.5*	67.9*
Other	80.9	68.7	73.6	38.5*	79.0	75.1
Total	78.2	**64.2**	71.3	45.0*	74.7	72.0

* Cells with fewer than 20 persons in the initial sample.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

These tables show that the response rates for most of the cells are similar to the overall response rate of 72.9 percent. Table 4.4 lists the five cells in which response rates were lower.

Table 4.4: **Cells with low response rates**

Cell	Initial sample size (% of total sample)	Response rate
Age 16 - 35 Unemployed, some secondary education	108 (2.1)	64.8
Age 16 - 35 Unemployed, completed post-secondary education	100 (2.0)	63.0
Age 36 - 55 Neither employed nor unemployed, some secondary education	103 (2.0)	65.0
Age 56 - 65 Employed, some secondary education	41 (0.8)	51.2
Age 56 - 65 Some secondary education	148 (2.9)	64.2

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

In examining the characteristics of these cells with lower response rates, it appears that people at either the low end of the literacy scale (older, did not complete secondary education) or the high end (younger, completed postsecondary education) were less likely than others to respond to the survey. In most cases, however, the response rates for these groups were not drastically different from the rate for the overall population. Although the 51 percent response rate for employed adults aged 56 to 65 with some secondary education is quite low, the initial sample size in this cell (41) is too small to have much of an impact on the findings.

Stage 2

In any survey, re-weighting for non-response is based on an implicit model of the non-response process. In the case of the IALS, the model assumed that while response rates may vary across demographic groups, respondents and non-respondents within each group have comparable literacy skills. Thus, within each cell, the sample weights associated with the non-respondents were redistributed evenly across the respondents. The Canadian weights were adjusted to province, economic region, metropolitan area, age, sex, in-school youth, out-of-school youth, unemployment insurance recipients, and social assistance recipients Census counts.

This phase of the analysis examined non-response bias *under this model of non-response*. The goal of the analysis was to estimate the extent of the bias in estimating the literacy profiles that is "fixed" by adjusting the sample weights. This approach is equivalent to one that would impute literacy scores for respondents under the same model. For example, cell-mean imputation or hot-deck imputation, in which the cells are defined by the same variables used in the weight adjustments, would produce substantially the same literacy profile results. The approach used here is a more efficient way to obtain the same information.

Table 4.5 presents Canada's literacy profile on the prose, quantitative, and document scales, as published in *Literacy, Economy and Society*. The table includes the 95 percent confidence intervals for the estimates.

Table 4.5: **Canada's literacy profile based on the main and special samples, with 95 percent confidence intervals**

	Prose		Document		Quantitative	
Literacy level	%	95% CI	%	95% CI	%	95% CI
1	16.6	(14.6, 18.6)	18.2	(14.4, 22.0)	16.9	(13.3, 20.5)
2	25.6	(22.0, 29.2)	24.7	(21.7, 27.7)	26.1	(21.1, 31.1)
3	35.1	(30.3, 39.9)	32.1	(28.5, 35.7)	34.8	(30.6, 39.0)
4	22.7	(18.1, 27.3)	25.0	(22.4, 27.6)	22.2	(18.6, 25.8)

Note: These data are for adults aged 16 to 65 from the main and special samples.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Estimates produced using only the main Canadian sample (selected from the Labour Force Survey frame) and the final IALS weights are shown in Table 4.6. These estimates are virtually the same as those presented in the previous table.

Table 4.6: **Canada's literacy profile based on the main sample only**

Literacy level	Prose	Document	Quantitative
1	16.5	18.2	16.9
2	25.5	24.7	25.9
3	35.1	32.1	34.9
4	22.9	25.3	22.3

Note: These data are for adults aged 16 to 65.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Assuming that the confidence intervals shown in Table 4.5 are also appropriate for the estimates in Table 4.6 (an assumption necessitated by cost factors), researchers compared the latter profile to the Canadian literacy profile estimates using the design weight (i.e., unadjusted for non-response). Table 4.7 presents this literacy profile with the confidence intervals of the final estimates.

Table 4.7: **Canada's literacy profile based on the main sample only, using the design weight, compared with the 95 percent confidence intervals using the final weights**

	Prose		Document		Quantitative	
Literacy level	Design weight %	Final weight 95% CI	Design weight %	Final weight 95% CI	Design weight %	Final weight 95% CI
1	14.7	(14.6, 18.6)	16.6	(14.4, 22.0)	14.6	(13.3, 20.5)
2	29.3	(22.0, 29.2)	25.6	(21.7, 27.7)	31.2	(21.1, 31.1)
3	39.4	(30.3, 39.9)	35.1	(28.5, 35.7)	40.0	(30.6, 39.0)
4	**16.6**	(18.1, 27.3)	22.7	(22.4, 27.6)	**14.2**	(18.6, 25.8)

Note: These data are for adults aged 16 to 65.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The effect of the weighting adjustment is to flatten the literacy level distribution. In other words, after the adjustment, larger percentages of the population score at the low and high ends of the scale, and correspondingly smaller percentages score in the middle two levels. This result implies that most of the non-respondents are at either the highest or the lowest levels of literacy.

It is important to note that the 95 percent confidence interval is a conservative estimate and does not take into account the variance of the estimate based on the design weight. Thus, if the weighting adjustments were not performed, the bias that would have occurred would have been in the level 4 estimate. On the prose and quantitative scales, fewer persons scored in level 4 before the weighting adjustments were performed.

The bias revealed by this analysis is not particularly damaging to the Canadian data. Because Canada (like the other IALS countries) adjusted its samples to known population counts, the downward bias in the unadjusted estimates in prose and quantitative levels 4/5 may have been accounted for. In other words, the adjustments needed to alleviate the bias, given the assumptions of the model, were made. Countries that did not benchmark their data as Canada did, however, may have slightly underestimated the number of persons in the higher literacy levels. Of course, this would occur only if other countries' non-response patterns were identical to the Canadian pattern.

Stage 3

The next step of the analysis approached the question of non-response bias from the opposite direction. The goal here was to estimate the level of bias that would be unacceptable. More specifically, the question is: what magnitude of difference in literacy skill profiles between non-respondents and respondents would be required in order to significantly change the country's mean literacy results?

A difference of 10 points (which represents 20 percent of the standard deviation) in mean scores would be considered a significant change in Canada's average literacy scores. Researchers therefore sought to determine what non-respondents' literacy levels would need to be in order to effect such a drop. These estimates could then be compared with the literacy proficiencies of respondents in the same age, education, and employment groups to see the extent to which respondents and non-respondents differed.

The following steps were required to carry out this phase of the study for each of the age, education, and employment groups:

1) Calculate the percent of the population represented by the non-respondents using the base (unadjusted) weights.

2) Calculate the mean score difference between non-respondents and respondents that is needed to achieve an overall score difference of 10 points. Suppose that the weighted response rate is p with a mean score of :. Then, for the mean to drop by 10 points, we must calculate b such that:

$$p(\mu) + (1 - p)(\mu - b) = : - 10$$

Solving for b, we have:

$$b = 10 / (1 - p)$$

In total, the design weights for Canadian respondents (*p*) represented 81.3 percent of the population. The mean scores of respondents and non-respondents would have had to have differed by 53.3 points in order for the Canadian overall mean to drop 10 percent. For example, on the document scale, to lower the current mean of 279.3 by 10 percent, to 269.3, non-respondents would have had to have scored 53.3 points below respondents on this scale. Such a drastic difference between respondents and non-respondents is unlikely. Such a difference would be equivalent to dropping a respondent from his current level of education to the next lower.

Table 4.7, which appeared earlier, supports this finding. It showed that the Canadian non-response adjustment served to increase the percentages of persons in the higher and lower literacy skill levels. Thus, to reduce the mean by 10 percent, those in the lower literacy skill levels would have had to have been much less skilled in order to effect this drop.

Although the same analysis was performed for each cell defined by age, education, and employment, the results of these analyses are difficult to interpret. Because the Canadian survey used a relatively large number of post-stratification cells, and these post-strata were different from and hence crossed the age/education/employment cells, the results contain much confounded information. For example, response rates of more than 100 percent were obtained for some cells. Hence, this part of the analysis is not reported.

4.3 United States Study of Non-response Bias

The United States used the Current Population Survey (CPS) in March 1994 as its sampling frame for the IALS. Because the CPS contained substantial background information about respondents, it is possible to compare the characteristics of IALS respondents and non-respondents and to investigate the likely difference in literacy skill profiles between the two groups.[2]

Researchers investigated non-response bias in the United States IALS in four stages:

Stage 1: Compare the characteristics of respondents and non-respondents for differences in response probabilities across demographic groups;

Stage 2: Compare estimates from the sample using the base weights with estimates derived using the final weights after post-stratification and adjustment for non-response;

Stage 3: Compare results using the sample data and the final, adjusted weights with estimates using weights based on different models of non-response; and

Stage 4: Establish an acceptable level of potential bias, and calculate the size of the average difference between respondents and non-respondents that would be required to cause bias of this magnitude.

The findings from these analyses were then synthesized to form a judgment about the likely biases due to non-response and the consistency of this bias across the United States and Canadian data. This study is restricted to those aged 16 to 65 since this was the age range of the IALS target population.

---

[2] The CPS achieved a response rate of 90 percent. The Census adjusted the CPS weights to account for non-response. This paper investigates bias as the difference between the achieved IALS sample and the achieved CPS sample, after weight adjustments.

Stage 1

According to the sampling plan, 4,875 people aged 16 to 65 were selected for the IALS in the United States; of these, 3,035 individuals responded, resulting in an overall response rate of 62 percent.

The investigation of response bias began by examining differential response rates across groups defined by age, education, employment status, sex, and race. Researchers examined response rates separately for each of the five demographic characteristics, and for groups defined by combining these demographic characteristics.

Table 4.8 presents the sample size, number of respondents, and response rate for each of the five demographic groups mentioned earlier. These data show that older people were more likely than younger people to respond to the IALS: 73 percent of those aged 56 to 65 and 67 percent of those aged 36 to 55 were respondents, compared with 55 percent of those aged 16 to 35. In terms of educational level, people with the least and the most education were more likely than others to respond. Response rates ranged from 59 percent for individuals with some postsecondary education to 72 percent for those with primary or less schooling. Similarly, response rates ranged from 54 percent for American Indians to 68 percent for Asian and Pacific Islanders. The response rates for American Indians, Blacks, and Hispanics were below the overall average. On the other hand, response rates varied little by employment status and sex.

Table 4.8: **Response rates by demographic characteristics**

	Number in CPS sample	Number of IALS respondents	Response rate (%)
Total	**4,875**	**3,035**	**62**
Age group			
16 to 35	2,308	1,234	55
36 to 55	1,951	1,281	67
56 to 65	616	478	73
Educational level			
Primary school or less	375	434	72
Lower secondary	1,151	435	61
Upper secondary	1,312	1,097	60
Some post-secondary	977	347	59
University degree	1,060	710	68
Employment status			
Employed	3,298	2,043	63
Unemployed	247	150	60
Other	1,330	775	62
Sex			
Male	2,359	1,433	61
Female	2,516	1,601	64
Race			
White	2,958	1,876	63
Black	850	489	58
Hispanic	696	421	60
American Indian	26	14	54
Asian/Pacific Islanders	336	228	68
Other	--	--	--

-- Indicates that the cell size is less than 20.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

IALS analyses have shown that age, educational level, and employment status are correlated with literacy level. Accordingly, the next analysis examined these three variables combined. Crossing these categories results in 3 x 5 x 3 = 45 comparison groups. The response rates within these 30 categories revealed little that was not apparent from Table 4.8. The detailed table is presented in Appendix I, Table I.1.

Stage 2

The adjustments to the weights originally implemented by the Census bureau were intended to offset the impact of differential response across race and educational groups. This section reviews the impact of these adjustments.

Table 4.9 contrasts estimates of the distribution of the population across the literacy levels before and after adjustment of the weights. The difference between the two figures, presented in Column 3, provides a measure of the bias fixed by non-response, assuming that the model of non-response is accurate. Column 4 presents an index of the size of the bias relative to other sources of error in the survey. Specifically, the standard error calculation presented in this fourth column incorporates both sampling error and measurement error of the literacy test. Column 5 presents the ratio of non-response bias to these other sources of error. For the most part, these ratios are moderate.

Overall, the literacy profile obtained using sample base weights is similar to that using adjusted final weights. The difference between the two profiles is approximately one standard error in only two cases: prose literacy levels 4 and 5 combined, and quantitative literacy levels 4 and 5 combined.

A similar pattern is evident among more detailed demographic groups (see Appendix I, Table I.2 through I.6). The differences obtained from using the base and final weights exceed one standard error for only five groups: employed, quantitative levels 4/5; males, quantitative levels 4/5; lower secondary education, prose level 1; White, quantitative levels 4/5; and Black, quantitative level 1. Recall that the smaller sample sizes within demographic groups lead to relatively large sampling error, however. Biases tend to be small in relation to the sampling error.

Table 4.9: **Comparing results based on base and final weights**

Scale/level	Base weight (1)	Final weight (2)	Difference: (1)-(2) (3)	Standard error (4)	(3)/(4) (5)
Prose					
1	20.5	20.7	-0.2	0.95	-0.21
2	24.9	25.9	-1.0	1.42	-0.70
3	32.2	32.4	-0.2	1.39	-0.14
4 and 5	22.4	21.1	1.3	1.33	0.98
Document					
1	23.6	23.7	-0.1	0.83	-0.12
2	25.0	25.9	-0.9	1.34	-0.67
3	31.5	31.4	0.1	1.17	0.09
4 and 5	19.9	19.0	0.9	1.35	0.67
Quantitative					
1	20.7	21.0	0.3	0.86	0.35
2	24.4	25.3	-0.9	1.48	-0.61
3	31.1	31.3	-0.2	0.96	-0.21
4 and 5	23.8	22.5	1.3	1.19	1.09

Note: The standard errors take into account sampling and measurement errors.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Under the model on which the reweighting is based, the biases from non-response are small—typically within one percentage point.

Stage 3

A more difficult realm of non-response analysis involves investigating the adequacy of the model of non-response that underlies the adjustments to the weights. Ultimately, one can never know whether the right model has been selected. Indeed, it is sensible to begin by assuming that the received model is wrong, even though it may be adequate. Here, a model is deemed adequate if it yields results that are substantially similar to competing, reasonable models.

Available time and resources limited this aspect of the investigation to one model. This model assumes that all of the relevant difference between respondents and non-respondents is captured by variables measured on the sample frames. It should be noted that other models can also estimate bias from unmeasured sources, given the right data.

For this analysis researchers expanded the variables admitted as possible contributors to non-response. As noted earlier, the IALS in the United States drew on a sample frame that is rich in ancillary information, since respondents had earlier completed the CPS questionnaires. The problem in incorporating these data into a non-response model using traditional weighting methods is that the cell sizes would quickly dissipate.

Rather than divide the sample into tiny cells and ignore many variables that might be important, researchers estimated a model in which the response probabilities are specified as a function of a linear combination of a moderately large set of predictor variables. In its simplest form, this model could consist of a linear regression of a binary dependent variable (response/no response) against the set of predictors. Perhaps the most important limitation of this linear probability model is that it can yield estimated probabilities outside the zero-one interval.

The probit model offers a more parsimonious alternative (see Greene 1990, 662-72, or Maddala 1983, 22-27). This model transforms a linear combination of predictor variables into a normal ogive, a form that approaches zero and one.

Non-response adjustments based on a probit response model were developed in three steps:

- Identify the variables to be included in the response model;
- Estimate the response model; and
- Estimate the non-response adjustment as the inverse of the predicted values from the response model.

The logic of this adjustment procedure is similar to that of the more traditional approach. If someone has a one in five probability of responding, then only one in five sampled people like him or her would respond. This individual, then, must represent five times the number of people he was originally intended to represent. The difference between this approach and the traditional approach to post-stratification is in the way we define "people like him or her" and the number of factors that can be included in this definition of similarity. The CPS data contain many candidate factors for inclusion in this model, including information about labor force participation, age in number of years, marital status, home ownership status, and urbanicity.

Estimates from the probit model are presented in Appendix I, Tables I.7 and I.8. Although time limitations prevented an in-depth analysis of the competing models, a simple test provides insight into the likely existence of substantive differences between competing models. The original non-response adjustments were done within the cells formed by crossing race and educational attainment (with a few cells collapsed together). If the probit model suggested the same adjustments, then the adjustment factors should be fairly constant within cells, and should vary across cells. An analysis of variance of the adjustments from the probit model found that just over 20 percent of the variance was between cells and the remainder was within cells. This suggests that the competing model could yield substantially different results than did the original non-interview adjustments.

Further research is required to establish the extent to which specific findings would differ under competing models of non-response. However, this brief investigation indicates that the original two-factor model used for the United States could be improved. Most notably, home ownership status predicted response probability. This result seems reasonable, because home owners are likely to be less mobile and more affluent than their non-owning counterparts. Given that little effort was made to track sampled individuals who had changed addresses, we ought to expect that home ownership would predict response probability. To the extent that home ownership is associated with literacy, it could produce bias.

The findings from this analysis suggest the presence of response bias that is not corrected by the original adjustment procedures. The weight adjustments suggested by the expanded model vary considerably from the original adjustments. The most important predictor of response in this model, home ownership, is likely to be correlated with literacy.

Stage 4

Thus far, the analyses have focused on estimating the likely bias from non-response. These analyses can be put in perspective by focusing on the country rankings and asking how big a bias would be required to alter our policy-relevant inferences from the data.

Researchers calculated the average difference between respondents and non-respondents that would be required to change the United States means or level percentages by 3, 5, 10, and 15 points. In most cases, a difference in the range of 6 to 10 percentage points would be required. Comparisons within subgroups are sometimes closer, however, so smaller biases might make a difference.

The steps required to carry out this analysis for each of the age, education, employment, sex, and race groups in the United States are as follows:

- Calculate the percent of the population represented by the non-respondents, r, using the base (unadjusted) weights, and
- Calculate the mean differences needed to achieve a difference of x points, which is

$(1-r)(u)+r(u-b)=u-x$ u = *mean score*

thus, $b = x / r$.

Table 4.10 presents these results. The table shows that non-respondents represent 37 percent of the United States population using the design weight. The difference in mean literacy scores between respondents and non-respondents would have to be 8.11

BEST COPY AVAILABLE

points in order to change the United States overall mean by 3 points, 13.51 points to change the overall mean by 5 points, 27.03 points to change the overall mean by 10 points, and 40.54 points to change the overall mean by 15 points.

Table 4.10: **Proportion of population represented by non-respondents and mean differences needed to achieve a difference of 3, 5, 10, and 15 points in United States literacy estimates**

	% of population represented by non-respondents	Average difference between respondents and non-respondents needed to bias country estimate by ...			
		3 points	5 points	10 points	15 points
Total	**37**	**8.11**	**13.51**	**27.03**	**40.54**
Age group					
16 to 35	44	6.82	11.36	22.73	34.09
36 to 55	32	9.38	15.63	31.25	46.88
56 to 65	25	12.00	20.00	40.00	60.00
Educational level					
Primary school or less	25	12.00	20.00	40.00	60.00
Lower secondary	41	7.32	12.20	24.39	36.59
Upper secondary	39	7.69	12.82	25.64	38.46
Some post-secondary	40	7.50	12.50	25.00	37.50
University degree	30	10.00	16.67	33.33	50.00
Employment status					
Employed	37	8.11	13.51	27.03	40.54
Unemployed	42	7.14	11.90	23.81	35.71
Other	37	8.11	13.51	27.03	40.54
Sex					
Male	39	7.69	12.82	25.64	38.46
Female	35	8.57	14.29	28.57	42.86
Race					
White	36	8.33	13.89	27.78	41.67
Black	42	7.14	11.90	23.81	35.71
Hispanic	37	8.11	13.51	27.03	40.54
American Indian	40	7.50	12.50	25.00	37.50
Asian/Pacific Islander	34	8.82	14.71	29.41	44.12
Other	--	--	--	--	--

-- Indicates that both the unweighted numerator and denominator are less than 20.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

For subgroups with lower response rates, somewhat smaller score differences between respondents and non-respondents would be needed to change the United States mean. For example, to achieve a 10 point change in overall mean scores for the United States, the mean score of non-respondents aged 16 to 35 (a group with a relatively low response rate) would need to be 22.73 points below that of respondents. For the 56 to 65 age group (higher response rate), a difference of 40 points would be required to achieve a similar impact.

In summary, non-response does appear to have introduced biases into the United States estimates for the IALS. Furthermore, bias probably remains after the non-interview adjustments, and these biases may be as large or larger than the combined sampling and measurement variance of the literacy tests in some cases.

This bias is unlikely to be large enough to effect the key inferences from the survey, however. The researchers found no evidence, for example, that the biases are large enough to alter the country rankings for key statistics such as the percent of the total population in particular literacy levels or the total population mean scores.

4.4 Swedish Study of Non-response Bias

For the IALS, a sample of 5,275 persons was drawn from Sweden's current population register of citizens over age 15. Of this sample, 167 persons were identified as not belonging to the target population, leaving 5,108 individuals, 3,038 of whom were interviewed. The response rate was 59.48 percent. The age range was adjusted at a later stage so as to make the sample conform to the international target population of adults aged 16 to 65. In all, 2,645 persons in the 16 to 65 age group took part in the Swedish IALS.

A stratified sampling procedure was used in which all municipalities were grouped into one of two categories: urban or rural. The first stratum contained 30 municipalities and the population register, complemented with cases drawn in two stages for certain suburban areas. A two-stage sample was also drawn for the second category, with municipalities chosen from each stratum and individuals drawn in numbers proportional to population size.

Evidence of representativity is presented in Figures 4.1 through 4.4, which compare the achieved sample with the population in terms of four variables: age, sex, country of birth, and foreign-born population by age. Table 4.11 presents data on the magnitude of error in the estimates for the total population of nonrespondents, and for a variety of subpopulations of nonrespondents, using a 95 percent confidence interval.

Figure 4.1: **Comparison of sample estimates to population counts, by age**

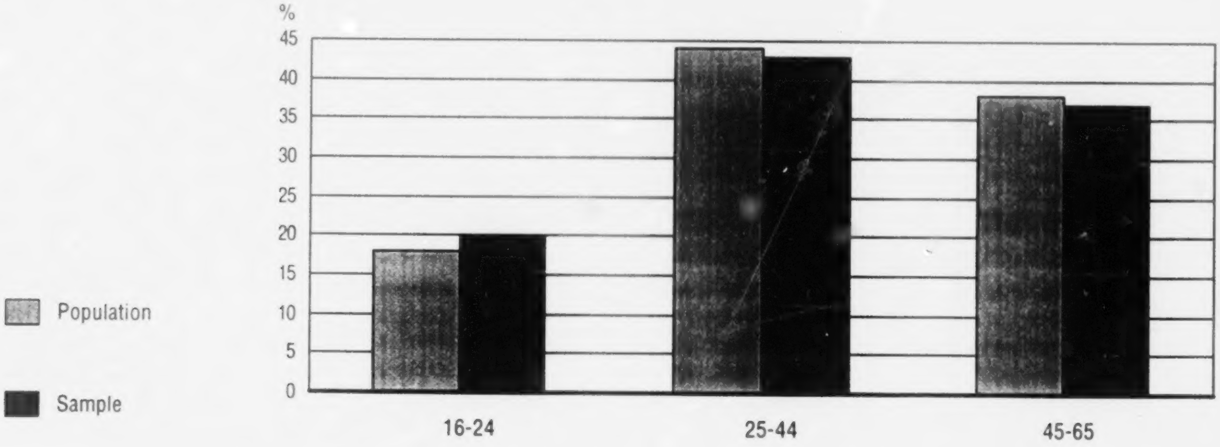


Figure 4.2: Comparison of sample estimates to population counts, by sex

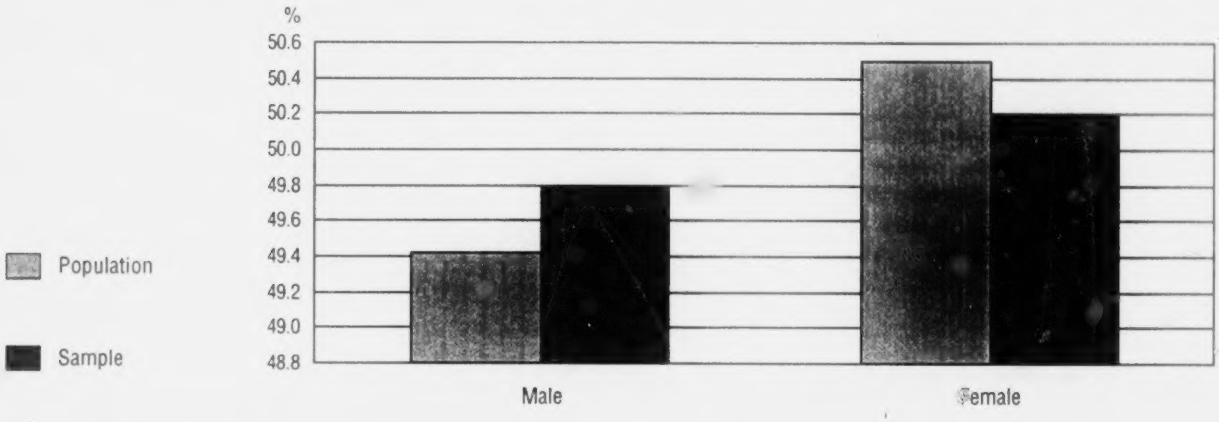


Figure 4.3: Comparison of sample estimates to population counts, by country of birth

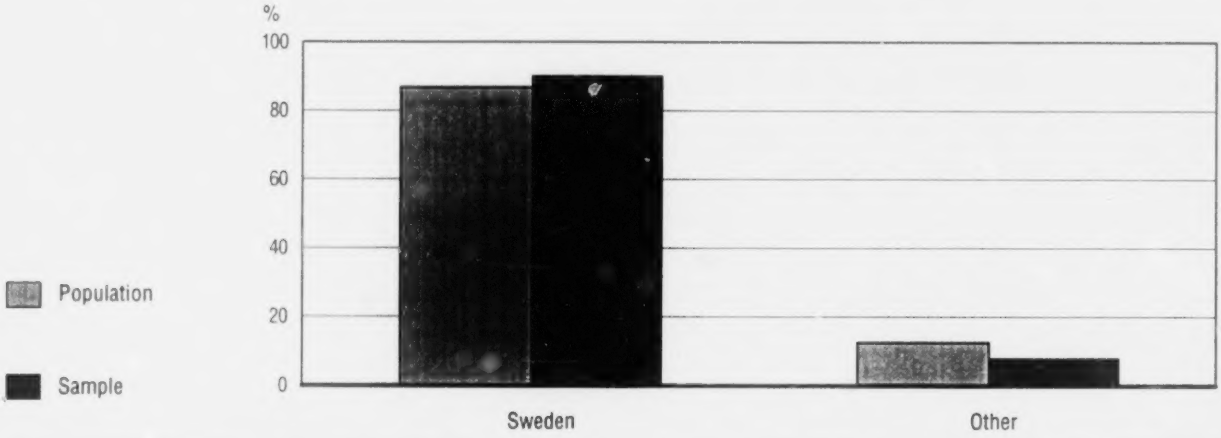


Figure 4.4: **Comparison of sample estimates to population counts of foreign-born, by age**

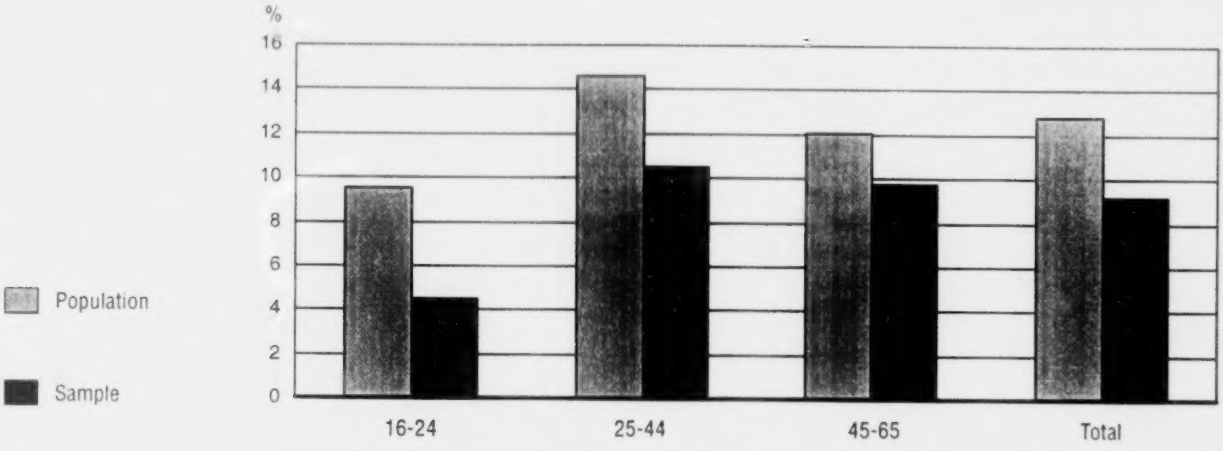


Table 4.11: **Sample sizes and 5 percent error margins compared with population sizes with 50 percent non-response***

Group	Number of respondents	Error estimates	
		5%	50%
Total	**2,645**	**0.8%**	**1.9%**
Prose			
Level 1	185	3.1	7.2
Level 2	516	1.9	4.3
Level 3	1,043	1.3	3.0
Level 4	688	1.6	3.7
Level 5	168	3.3	7.6
Level 4/5	856	1.5	3.3
Document			
Level 1	156	3.4	7.8
Level 2	477	2.0	4.5
Level 3	1,027	1.3	3.1
Level 4	732	1.6	3.6
Level 5	203	3.0	6.9
Level 4/5	935	1.4	3.2
Quantitative			
Level 1	163	3.3	7.7
Level 2	472	2.0	4.5
Level 3	1,011	1.3	3.1
Level 4	719	1.6	3.7
Level 5	221	2.9	6.6
Level 4/5	940	1.4	3.2
Sex			
Male	1,289	1.2	2.7
Female	1,356	1.2	2.7

* Error calculated with assumed n<0.1 x population sizes.

Table 4.11: **Sample sizes and 5 percent error margins compared with population sizes with 50 percent non-response*** (Concluded)

Group	Number of non-respondents	Error estimates 5%	50%
Country of birth			
Sweden	24	0.9	2.0
Other	241	2.8	6.3
Age			
16 to 20	267	2.6	6.0
21 to 30	597	1.7	4.0
31 to 40	545	1.8	4.2
41 to 50	583	1.8	4.1
51 to 65	653	1.7	3.8
Educational attainment			
Tertiary, ≥3 years	635	1.7	3.9
Tertiary, <3 years	338	2.3	5.3
University	297	2.5	5.7
At most elementary	639	1.7	3.9
Elementary and lower secondary	310	2.4	5.6
Upper secondary (theoretical/vocational)	971	1.4	3.1
Higher education	1,214	1.2	2.8
Folk school, ≤8 years	322	2.4	5.5
First language			
Swedish	2,381	0.9	2.0
Finnish	69	5.1	11.8
Other	195	3.1	7.0
Labor market status			
Unemployed	217	2.9	6.7
Student	359	2.3	5.2
Employed	1,600	1.1	2.5
Self employed	145	3.5	8.1
Retired	154	3.4	7.9
Long-term illness or early retired	144	3.5	8.1
Age x education			
Age 16 to 35			
At most lower secondary	158	3.4	7.8
At most upper secondary	601	1.7	4.0
Tertiary	214	2.9	6.7
Age 35 to 55			
At most lower secondary	305	2.4	5.6
At most upper secondary	314	2.4	5.5
Tertiary	342	2.3	5.3
Age 56 to 65			
At most lower secondary	176	3.2	7.4
At most upper secondary	56	5.7	13.1
Tertiary	79	4.8	11.0

* Error calculated with assumed n<0.1 x population sizes.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

There is no evidence of systematic or significant differences between respondents and non-respondents in Sweden's IALS sample. After examining the survey data and reviewing the study design, the Swedish expert panel did not find any quality defect that would preclude publishing the country's data.

To comply with national standards regarding data quality, however, the Swedish National Study Team commissioned a non-response follow-up study. The preliminary findings of this study are summarized below.

Based on the whole sample of 3,038 persons, the non-response rate for Sweden was about 40 percent. Of these non-respondents, 21.5 percent could not be located, 54.5 percent refused to take the test but answered some of the questions in the background questionnaire, and 16.8 percent refused to respond to both the test and the background questionnaire. The remaining 7.2 percent did not participate in the survey due to a handicap or illness, or for some other reason.

A non-response follow-up study was undertaken to learn more about the non-respondents and their likely proficiency levels. A sample of 150 persons was drawn randomly from Sweden's non-respondents. Of these, 78 persons (52 percent) participated in the follow-up study, in which they were asked to respond to the IALS instruments.

The 78 participants in the non-response study can be divided into two groups. The first group consists of 47 persons who took the IALS instruments under the same conditions as those in the main survey. The results were comparable to those obtained in the main survey. The second group consists of 31 persons who took the IALS instruments alone in their homes at a time they chose. The results for this group—in which men, the middle aged, and immigrants are over-represented—are much better than those for the comparison group. The analysis shows that the results obtained by the 31 persons who took the test without the presence of an interviewer are not reliable.

4.5 Conclusions

The results of the analyses described in this chapter indicate that the magnitude of non-response bias in the Canadian and Swedish IALS surveys was small.

The Canadian study suggests that non-response affected the percentages of persons at the highest and lowest literacy skill levels, and that this would have been important had a benchmarking adjustment not been performed. Even in other countries which did not benchmark in the same manner, however, the proficiency differences between respondents and non-respondents would have to be very large to produce a noticeable change in the country's literacy skill profile.

Non-response does appear to have biased the IALS estimates for the United States, even after the non-interview adjustments were performed. Although this bias may be as large or larger than the combined sampling and measurement variance of the literacy tests in some cases, it is unlikely to be large enough to affect the survey's major findings and inferences.

It is important for future surveys to learn from the IALS experience. Although the magnitude of bias introduced into the estimates appears to have been small, and hence of little consequence, serious problems could have occurred. We strongly recommend that countries participating in future rounds of the IALS rely on data collection organizations which are experienced in fielding large-scale, high-quality household surveys. National study teams must also be provided with sufficient financial resources to ensure the integrity of the work. Further, the non-response analysis conducted for the IALS was not as extensive or as definitive as desired. Future studies should be built with such an analysis in mind right at their planning stages.

Part II:

Non-sampling Error

Introduction to Part II

As Part I makes clear, the IALS study teams devoted considerable effort to ensuring that the survey samples were high-quality scientific probability samples capable of supporting international comparisons of adult literacy. Even the best of sampling plans can be implemented improperly, however. The chapters in this part of the report present evidence related to the presence and impact of non-sampling error in the IALS estimates.

Chapter 5 presents information about data collection, scoring, and data processing, while Chapter 6 explores the use of incentives and discusses respondents' motivation to participate and to perform well in the IALS.

Chapter 5:

Data Collection and Processing

Nancy Darcovich and T. Scott Murray

5.1 Introduction

The IALS gathered descriptive and proficiency information from sampled respondents through a background questionnaire and a series of assessment booklets containing prose, document, and quantitative literacy tasks. Survey respondents spent approximately 20 minutes answering a common set of background questions concerning their demographic characteristics, educational experiences, labor market experiences, and literacy related activities. Responses to these background questions make it possible to summarize the survey results using an array of descriptive variables, and also increase the accuracy of the proficiency estimates for various subpopulations, as described later in this report. Background information was collected by trained interviewers.

After answering the background questions, the remainder of respondents' time was spent completing a booklet of literacy tasks designed to measure their prose, document, and quantitative skills. Most of these tasks were open-ended; that is, they required respondents to provide a written answer.

To achieve good content coverage of each of three literacy domains, the number of tasks in the assessment had to be quite large. Yet, the time burden for each respondent also needed to be kept within an acceptable range. To accommodate these two conflicting requirements—in other words, to reduce respondents' time burden without sacrificing good representation of the content domain—each respondent was administered only a fraction of the pool of tasks, using a variant of matrix sampling.

The success of the IALS depended first and foremost on the development and standardized application of a common set of survey instruments. Although each

IALS country designed and implemented its own data collection activities, numerous procedures were implemented to secure the quality of this part of the study, and of the scoring and data processing procedures. For example, participating countries were given model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments and for handling non-response codings. These and other actions concerning IALS data collection, scoring, and data processing are described in the remaining sections of this chapter.

5.2 Model Procedures Manuals and Instruments

Each IALS country was given a set of administration manuals and survey instruments to use as a model (see Appendix C). Countries were permitted to adapt these models to their own national data collection systems, but they were required to retain a number of key features. First, respondents were to complete the core and main test booklets alone, in their homes, without help from another person or from a calculator. Second, respondents were not to be given monetary incentives for participating. Third, despite the prohibition on monetary incentives, interviewers were provided with procedures to maximize the number of completed background questionnaires, and were to use a common set of coding specifications to deal with non-response. This last requirement is critical. Because non-completion of the core and main task booklets is correlated with ability, background information about non-respondents is needed in order to impute cognitive data for these persons.

Background questions

The model background questionnaires given to all IALS countries contained two sets of questions: mandatory questions, which all countries were required to include; and optional questions, which were recommended but not required. Countries were not required to field literal translations of the mandatory questions, but were asked to respect the conceptual intent of each question in adapting it for use. Countries were permitted to add questions to their background questionnaires if the additional burden on respondents would not reduce response rates.

Statistics Canada reviewed all background questionnaires except Sweden's before the pilot survey and offered comments and suggestions to each country.

Literacy tasks

The IALS is based on the premise that the difficulty of various literacy tasks is determined by certain factors which are stable across language and culture. Accordingly, all of the IALS countries were given graphic files containing the pool of IALS literacy items and were instructed to modify each item by translating the English text to their own language without altering the graphic representation.

Certain rules governed the item modification process. For instance, some items required respondents to perform a task that was facilitated by the use of keywords. In some cases, the keywords were identical in the question and the body of the item; in others, the keyword was similar but not exactly the same; and in still other cases, the keyword was a synonym of the word used in the body of the item. In another case, respondents were asked to choose among multiple keywords in the body of the item, only one of which was correct. Countries were required to preserve these conceptual associations during the translation process.

Particular conventions used in the items—for example, currency units, date formats, and decimal delimiters—were adapted as appropriate for each country.

To ensure that the adaptation process did not compromise the psychometric integrity of the items, each country's test booklets were carefully reviewed for errors of adaptation. Countries were required to correct all errors found.

The analysis presented in Chapter 9 of this report suggests that this review process was imperfect in two important respects. First, it is clear that countries chose not to incorporate a number of changes which were identified during the course of the review, believing that they "knew better". From a managerial perspective, little can be done about such cases, except to render explicit the inherent risks associated with these decisions. Second, the availability of empirical data from the study has permitted the identification of several additional sources of task and item difficulty which were not included in the original Kirsch-Mosenthal grammar. Item adaptation guidelines and item review procedures associated with subsequent rounds of IALS data collection have been adapted to reflect this additional information.

Standardized non-response coding

It was crucial that the IALS countries manage non-respondent cases in a uniform manner so as to limit the level of non-response bias in the resulting survey estimates.

In IALS, a respondent had to complete the background questionnaire, pass the core block of literacy tasks, and attempt at least five tasks per literacy scale in order for researchers to be able to estimate his or her literacy skills directly. Literacy proficiency data were imputed for individuals who failed or refused to perform the core literacy tasks and for those who passed the core block but did not attempt at least five tasks per literacy scale. Because the model used to impute literacy estimates for non-respondents relies on a full set of responses to the background questions, IALS countries were instructed to obtain at least a background questionnaire from sampled individuals. They were also given a detailed non-response classification to use in the survey. Chapter 11 describes the non-response classification scheme and the imputation procedures.

5.3 Field Operations

Each country was responsible for hiring its own interviewing staff. Thus, the number of interviewers, their pay rates, and the length of the survey period varied among the countries according to their norms and budgets. Each country was provided with a booklet to be used in training interviewers (see Appendix D).

Table 5.1 lists the number of interviewers in each IALS country and the average number of interviews per interviewer, and indicates whether or not an introductory letter was sent to sampled households before the interviewers initiated contact.

Table 5.1: **Data collection summary**

Country	Number of interviewers	Average number of response interviews per interviewer	Introductory letter sent
Canada	250	23	Yes
France	135	22	No
Germany	327	6	Yes
Netherlands	150	21	Yes
Poland	200	15	Yes
Sweden	100	30	Yes
Switzerland	300	10	Yes
United States	149	21	Yes

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 5.2 summarizes the interview payments across the IALS countries.

Table 5.2: **Interview payment by country (in $USD)**

Country	Payment for incomplete interview	Payment for background questionnaire only	Payment for complete interview
Canada	— — — — — — — — — — — — — — —	one hourly rate	— — — — — — — — — —
France	not available	9 - 15	26 - 31
Germany	0	0	38 - 42
Netherlands	3	18 - 24	18 - 24
Poland	0	4	7 - 8
Sweden	not available		
Switzerland	not available	21	42
United States	— — — — — — — — — — — — — — —	one hourly rate	— — — — — — — — — —

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Aside from Canada and the United States, which paid interviewers by the hour, the IALS countries paid interviewers by the interview, paying most for a fully completed survey and far less for a non-response. Although this payment schedule encourages interviewers to obtain as many complete interviews as possible, it may also cause them to follow up less thoroughly with non-respondents. As shown later in this report, however, the response rates for Canada and the United States are no higher than those for the other participating countries.

Information about the interviewers, data collection process, training, and data collection costs for each IALS country is provided below.

Canada

Data collection was carried out separately for the main survey and the Ontario-Francophone survey because the latter survey included a screening process to identify persons whose mother tongue was French.

Professional interviewers from the national Labour Force Survey were employed to conduct the main IALS survey in Canada. Professional interviewers as well as recently hired interviewers were employed to conduct the Franco-Ontarian survey. A total of approximately 250 interviewers across Canada worked on the survey for a three-week period. The average workload per interviewer was about 25 interviews.

Interviewers' training consisted of a 4-hour self-study package as well as a 1-1/2 day training course. Interviewers were paid by the hour, and a total of approximately $344,500USD was spent on their salaries.

Approximately 40 senior interviewers supervised the data collection to ensure that the proper procedures were being followed and that the questionnaires with status codes were being correctly completed. Direct feedback was provided to interviewers to correct problems that emerged during the data collection process.

France

The interviewers used in the French survey had solid survey experience: on average, they had worked 4.2 years for the data collection firm that administered the survey. Interviewers were trained in a one-day session which covered the presentation of the survey, the methodology used, a detailed study of the forms, administration of the tests, and information about how to proceed in difficult cases. Interviewers were also given detailed written instructions. In all, 135 interviewers worked on France's IALS, conducting an average of 22 interviews each. Field operations were carried out between 10 October 1994 and 25 December 1994.

Interviewers were paid 130 FF ($25USD) for each completed interview outside Paris and 155FF ($30USD) for each completed interview within Paris. If only the background questionnaire was completed, the interviewer was paid half of the fee. Interviewers who worked far from their homes received payment for travel time as well as for travel and accommodation expenses.

The French interviewers were supervised through checks of the completed forms. When errors were found, the supervisor either telephoned or visited the respondent. Fifteen percent of the interviews conducted by each interviewer were verified. In cases in which errors were found, the supervisor re-visited the respondent.

French interviewers conducted their interviews without prior contact with the selected household. At the initial contact, they provided a letter from the Department of Education.

Germany

In Germany, the firm that conducted the IALS pilot test also conducted the main survey. A group of 327 interviewers worked on the main survey, 19 of whom were assigned to one region of the country. On average, each interviewer conducted 6.7 interviews. The data were collected from 2 September 1994 to 12 November 1994.

Most of the interviewers were part-time workers, many of whom had a second job. Most had two to three years of experience and were required to have worked on at least 10 previous surveys.

The German interviewers were given no special training for the IALS survey, but all received general interviewer training. They were also given written instructions for IALS and instructed to call the head office in case of problems.

To supervise its interviewers, the company conducting the IALS verified the selected households through a postal check for every tenth survey conducted. Postal checks were also carried out if problems were observed in the completed questionnaires. If the response rate to the postal check was too low, a telephone follow-up was performed.

The interviewers in Germany were paid 55 to 60 DM ($35 to $38USD) for each completed interview. If a respondent did not complete the literacy test, the interviewer was not paid.

Each household selected for the sample received a letter advising members of the household that a survey would be carried out. This letter stated that one aim of the survey was to improve the quality of the newspapers, instructions, and forms written for the public. The letter also mentioned a lottery with 600 prizes for respondents, including a first prize of 500 DM ($320USD).

Interviewers introduced the IALS instruments by saying that one of the survey's aims was to improve the quality of newspapers, instructions, and forms written to the public, rather than by saying that the survey was actually intended to test the respondent's literacy level. Respondents were given the core tasks with the main booklet of tasks. Because the interviewers did not score the core tasks, they were not used as a screener. A small percentage (1.8 percent) of respondents failed the core tasks but went on to complete the main booklet.

Interviewers were instructed to check the time taken for each block. After 20 minutes, they were allowed to suggest that the respondent skip to the next block. On average, the interviews lasted 92 minutes.

Respondents were permitted to use a calculator to perform some of the literacy tasks, although the interviewer was not allowed to provide it. Calculator use was coded during the scoring process, and items on which respondents used a calculator were excluded from the estimation of proficiency values. Germany's results were therefore unlikely to have been affected by calculator use as only nine such cases occurred.

Netherlands

The Netherlands administered the IALS background questionnaire using Computer Assisted Interviewing (CAI). Professional interviewers with three to four years of experience were employed to administer the survey. Many of these interviewers had administered the pilot test as well.

Interviewers were trained in a one-day session in which the principles of the survey were explained and videos with mock interviews were shown. Most interviewers were unfamiliar with the core and main tasks booklets, so the training stressed the administration of these tests.

Approximately 150 interviewers worked on the survey. The workload per interviewer ranged from 20 to 120 cases, depending on the number of hours that the interviewer was scheduled to work. The interviews were conducted over a three to four month period, including the time allowed for follow-up.

Interviewers were paid $18 to $24USD per completed questionnaire, depending on the amount of travel involved. Interviewers were paid the full amount even if only part of the test or only the background questionnaire was completed. For flat refusals, $3USD was paid if at least the three required demographic variables (sex, age, and education) were obtained.

Eight supervisors oversaw the work of the interviewers. Four were trained specially on the IALS survey, and another two were familiar with field work in general. These supervisors verified the interviewers' work either by accompanying them to an interview or by verifying with respondents that contact had indeed been made. In addition, the CAI system was equipped with a bulletin board on which supervisors could post notes. This two-way communication system facilitated discussion between interviewers and supervisors.

In introducing the Dutch IALS survey, the interviewers read from a script shown on a computer screen. The script noted that the IALS was an international project, that it had some easy and some hard questions, and that the respondent should attempt to answer all questions, should use glasses if necessary, and should take the time necessary to complete the questionnaires.

The CAI component containing the background questionnaire had range edits and cross-question checks built into it. In the Netherlands, an eighth block was added to the end of the main tasks booklets. This block was the IEA reading test. During data entry of the test scores, each string of scores was 20 percent verified.

Data collection procedures in the Netherlands followed those as outlined in the guidelines with one exception. In administering the interviews, no interpreters were used to obtain background information from those who did not speak Dutch. Accordingly, no literacy profile could be constructed for non-Dutch speakers.

The Dutch addressed this issue in their fieldwork report, stating:

> "The Netherlands is officially a one language country. There are some dialects and there is one official regional minority language (Frisian) but research shows that people from Friesland have a similar mastery of the Dutch language as non-Frisians.
>
> As far as people are concerned who were born outside of the Netherlands or whose parents were born outside of the Netherlands, research shows that the second generation speaks mostly Dutch in the outside world (school, work, society) and the language of origin at home. The third generation speaks Dutch in (almost) all circumstances.
>
> Language difficulties are thus concentrated in the first generation, those that were born outside of the Netherlands. For the total population this amounts to 1,375,000 people (9% of the population).
>
> Parts of this group, and parts of the group of second generation children as well, might have less mastery of the Dutch language than those who were born here and whose parents were born here as well. However it is very difficult to estimate the severity of this problem. It might influence their score on the assessment, but it probably wouldn't prevent them from taking part in the assessment. If there is an influence much of it will be mediated through their level of educational attainment, for which the sample is weighted. Of the people not born in the Netherlands (no age restrictions) 55.2% had Dutch nationality. The biggest groups with Dutch nationality are formed by people from the former colonies, Indonesia (13.4% of all foreign born) Surinam (13.3%) and the Antilles with Aruba (4.7%). In many other cases, people born abroad with Dutch nationalities will be children of people working abroad (Belgium and Germany seem to be a case in point). It is probably safe to assume that most of the people in this group had a Dutch upbringing (Dutch as home language) or that Dutch was the language of instruction at school. The latter certainly applies to the former colonies like Indonesia and Surinam. Still, there are indications that language problems do arise in the group from

Surinam and the Antilles. But this would hardly prevent them from participation in the project."

Poland

In conducting the IALS, Poland employed professional interviewers, most of whom had about two years of survey experience. Interviewers were trained using a manual, a test interview, and a training session.

Approximately 200 interviewers worked on the survey with an average workload of 15 interviews per interviewer. The data collection activities occurred between October 1994 and January 1995.

Interviewers were paid by the questionnaire. No payment was received for a refusal. When only a background questionnaire was completed, the interviewer received about $4USD. When only the background questionnaire and core tasks booklet were completed, the interviewer received about $7USD. When all three instruments were completed, the interviewer received between $7 and $8USD, depending on the degree of completion of the main tasks booklet.

Approximately 10 supervisors worked on the survey. To control field procedures, they gave feedback directly to the interviewers. A judgment sample of 300 cases was drawn from the IALS sample and a second contact made with the respondent to check the procedures that had been used by the interviewer.

In analyzing the response codes from Poland's IALS instruments, it was found that close to 60 percent of those who completed a main tasks booklet did so only partially. The introduction to the survey mentioned problems with the quality of the main tasks booklet, and this may have led respondents to believe that the survey was a quality review of the main tasks booklet rather than a test of their abilities. The IALS study was designed to produce reliable estimates of the proficiency of population subgroups rather than individuals. Thus, the Polish introduction does not misrepresent the goals of the test and, in the professional judgement of the Polish National Study Team, minimised response bias which would otherwise have resulted. As a result, respondents may have lacked the motivation to complete the tasks.

Sweden

Approximately 100 interviewers conducted interviews for the Swedish IALS. Each interviewer was responsible for 40 to 50 addresses. Introductory letters describing the study and its aims were sent to respondents. Incentives (a "flower check" or a "movie check" worth about $7USD) were offered to those taking part in the survey.

Respondents who were immigrants with a limited command of Swedish were informed that the background questionnaire could be answered with the assistance of a "family interpreter." Respondents were not permitted to obtain extra help when performing the literacy tasks, however.

Switzerland

In Switzerland, about 300 interviewers conducted the IALS survey with an average of slightly more than eight interviews per interviewer. The interviewers were part-time workers with experience in at least one survey.

All Swiss interviewers received general survey training and also attended a three-hour IALS training session. Interviewers were asked to phone the training center at three different times during the survey process. The written instructions for the interviewers were not very detailed, but some instructions were given directly on the survey forms. To monitor the interviewers, follow-up was performed on 15 percent of the interviews.

Interviewers were paid 50FS ($40USD) for a fully completed survey. If only the background questionnaire was completed, the interviewer was paid half this amount.

In introducing the survey to respondents, the interviewers explained that the aim of the survey was to improve the quality of public documents.

United States

In the United States, Census Bureau interviewers were employed to conduct the IALS. All of them had at least five days of interviewer training. They were given a one-day training on IALS and were provided with substantial training and reference materials based on the Canadian training package. They also performed a day of field training under the supervision of a regional office supervisor.

In all, 149 interviewers were employed for the survey, and each had an average workload of 33 interviews. They were supervised by six regional supervisors who reviewed and commented on their work.

Before the data collection, a letter was sent to the selected addresses describing the upcoming survey. No incentive was offered. The survey was limited to 90 minutes. If a respondent took more than 20 minutes per block, the interviewer was instructed to move the respondent on to the next block. The United States interviewers were paid an hourly rate regardless of whether or not the questionnaires were completed.

5.4 Scoring

Respondents' literacy proficiencies were estimated based on their performance on the cognitive tasks administered in the assessment. Unlike multiple-choice questions, which are commonly used in large-scale surveys and which offer a fixed number of answer choices, open-ended items such as those used in the IALS elicit a large variety of responses. Because raw data is seldom useful by itself, responses must be grouped in some way in order to summarize the performance results. As they were scored, responses to the IALS open-ended items were classified as correct, incorrect, or omitted.

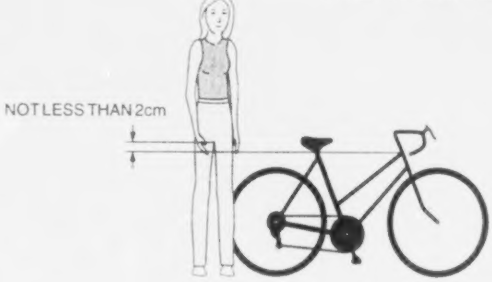
The models employed to estimate ability and difficulty are predicated on the assumption that the scoring rubrics developed for the assessment were applied in a consistent fashion within and between countries. Several steps were taken to ensure that this assumption was met.

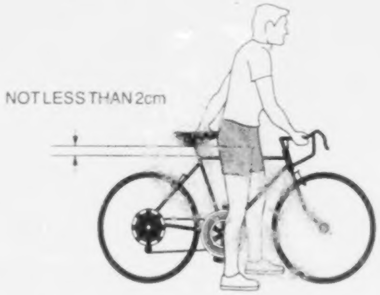
Analysis of pilot survey data

Analysis of data from the pilot survey resulted in the detection of numerous cases in which the scoring rubrics were being interpreted in an ambiguous fashion. This information was used to revise the scoring rubrics for those items retained for the main assessment.

PROPER FRAME FIT

RIDER MUST BE ABLE TO STRADDLE BICYCLE WITH AT LEAST 2 cm CLEARANCE ABOVE THE HORIZONTAL BAR WHEN STANDING.





NOTE: Measurement for a female should be determined using a men's model as a basis.

PROPER SIZE OF BICYCLE

FRAME SIZE	LEG LENGTH OF RIDER
430mm	660mm-760mm
460mm	690mm-790mm
480mm	710mm-790mm
530mm	760mm-840mm
560mm	790mm-860mm
580mm	810mm-890mm
635mm	860mm-940mm

OWNER'S RESPONSIBILITY

1. **Bicycle Selection and Purchase:** Make sure this bicycle fits the intended rider. Bicycles come in a variety of sizes. Personal adjustment of seat and handlebars is necessary to assure maximum safety and comfort. Bicycles come with a wide variety of equipment and accessories . . . make sure the rider can operate them.
2. **Assembly:** Carefully follow all assembly instructions. Make sure that all nuts, bolts and screws are securely tightened.
3. **Fitting the Bicycle:** To ride safely and comfortably, the bicycle must fit the rider. Check the seat position, adjusting it up or down so that with the sole of rider's foot on the pedal in its lowest position the rider's knee is slightly bent.

Note: Specific charts illustrated at left detail the proper method of determining the correct frame size.

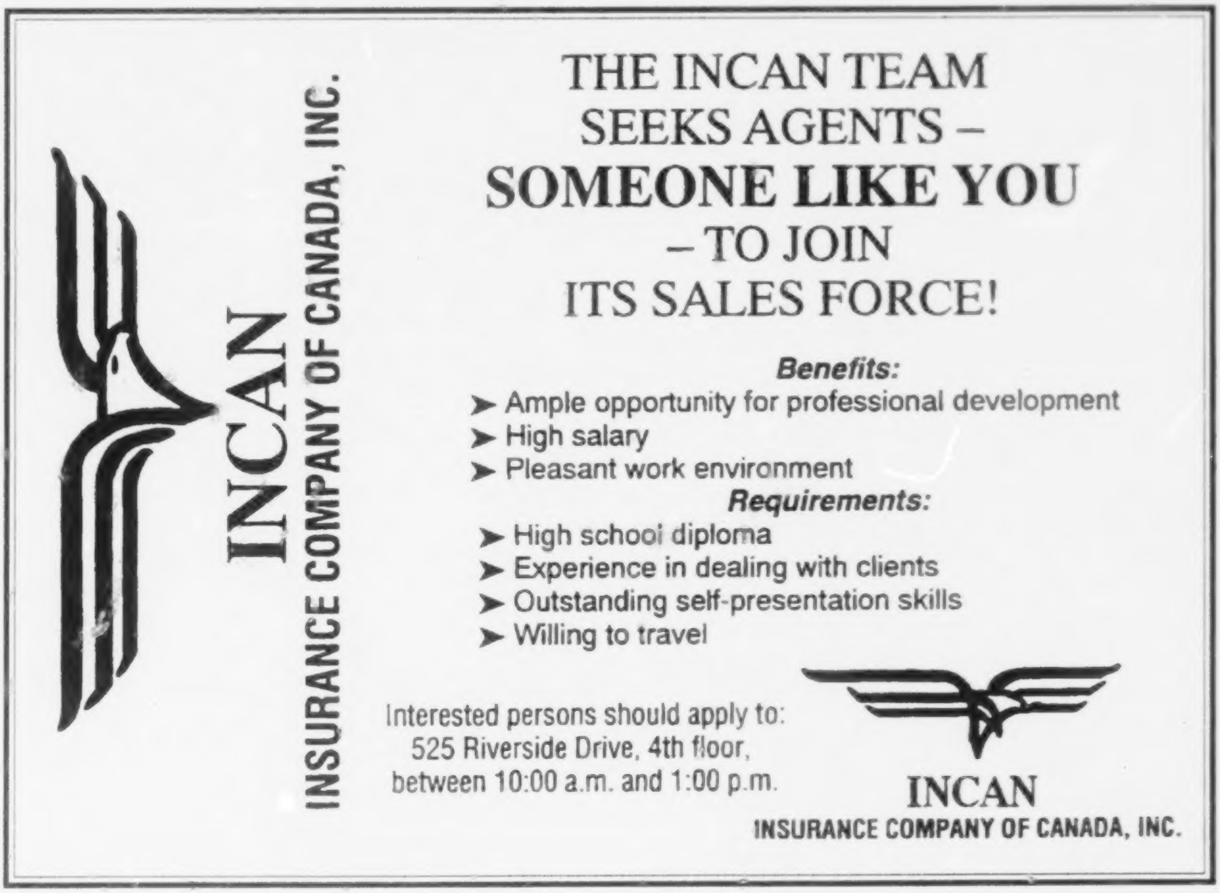
The manufacturer is not responsible for failure, injury, or damage caused by improper completion of assembly or improper maintenance after shipment.

For example, one task in the IALS pilot study gave respondents the following material and asked them, in English: "What size bicycle frame sizes should you consider if your legs are 810mm?"

The correct answer for this particular item was: 530mm, 560mm, and 580mm. This meant that respondents were required to include all possible frame sizes as part of the answer; any other response was to have been scored as incorrect (even if two out of the three sizes were listed by the respondent).

In the pilot study, the French-Canadian version of this item asked only for the "grandeur de cadre" or the size (singular), rather than the sizes (plural) asked for in the English version. Because the French-Canadian translation failed to articulate the plural nature of the correct response, widely divergent results were noted between the two populations on this particular item.

The French translation of this item also failed to ask for the plural "sizes," but France's scoring team accepted any one of the three sizes (530mm, 560mm or 580mm) as a correct response. This meant that while respondents in France typically provided only one of the three answers required for a correct score in other countries, they were not penalized with an incorrect score. Accordingly, for the main survey, the translations (in both French Canada and France) were corrected and the scoring guide was updated to clearly indicate that at least two of the three measures were needed for a correct answer.



INCAN
INSURANCE COMPANY OF CANADA, INC.

THE INCAN TEAM
SEEKS AGENTS –
SOMEONE LIKE YOU
– TO JOIN
ITS SALES FORCE!

Benefits:

- Ample opportunity for professional development
- High salary
- Pleasant work environment

Requirements:

- High school diploma
- Experience in dealing with clients
- Outstanding self-presentation skills
- Willing to travel

Interested persons should apply to:
525 Riverside Drive, 4th floor,
between 10:00 a.m. and 1:00 p.m.

INCAN
INSURANCE COMPANY OF CANADA, INC.

Another item in the pilot study was dropped from the IALS item pool as a result of its performance. In this case, respondents were shown the following advertisement and asked: "What job is advertised?" They were instructed to enter their answer on the line provided.

The correct answer for this item was: "(Insurance) Agent." Thus, "Insurance" was an optional part of the answer, but the word "Agent" was required. Unfortunately, many respondents focused on the "sales" component of the item and offered answers such as: "Insurance Salesman," "Insurance Salesperson" or simply "Salesperson." In Canada, these were scored as incorrect responses. In Germany and the Netherlands, however, the term used was simply "Sales Agent," all in one word, so the confusion between "Agent" and "Salesperson" never arose. Given that both responses are intuitively correct and that it would have been difficult to ensure that the item was functioning properly before the main study was conducted, this item was dropped from the survey.

It is important to note that for the pilot study, participating countries were required to blindly rescore all items and to resolve any scoring discrepancies identified. This step virtually eliminated scoring error in the pilot study, both within and between countries.

Quality control

To reinforce the importance of consistent scoring, a meeting of national study managers and chief scorers was held in Amsterdam prior to the commencement of scoring for the main study. The group spent two days reviewing the scoring rubrics for all the survey items. Where this review uncovered ambiguities and situations not covered by the guides, clarifications were agreed to collectively, and these clarifications were then incorporated into the final rubrics. By discussing the intent underlying each item in relation to the scoring rubric, chief scorers would be in a much better position to make judgments when confronted with unusual cases in scoring the main assessment.

To provide ongoing support during the scoring process, Statistics Canada and ETS maintained a joint scoring hotline. Any scoring problems encountered by chief scorers were resolved by this group, and decisions were forwarded to all national study managers. Together, these procedures give us confidence that the national study managers and chief scorers understood the scoring rubrics in detail going into the scoring exercise.

In addition to conducting intensive scoring training using the scoring manual, the study managers sought to improve the accuracy of scoring by discussing unusual responses with scorers and by offering additional training to some scorers, as needed, to raise their accuracy to the level achieved by other scorers.

To maintain coding quality within acceptable levels of error, each country undertook to rescore a minimum of 10 percent of all assessments. Roughly half of the rescored assessments were to be drawn from early returns so that scoring differences between scorers could be detected and corrected. The balance was to be randomly selected from the remaining assessments to guard against scoring drift. Where significant problems were encountered in either sample, larger samples of a particular scorer's work were to be reviewed and, where necessary, their entire assignments rescored. Countries were not required to resolve contradictory scores in the main survey (as they had been in the pilot), since outgoing agreement rates were far above minimum acceptable tolerances.

It is recognized that, despite the various quality assurance procedures implemented at the national level, significant differences could exist in the consistency of scoring between countries. To detect and estimate the magnitude of any such differences completely would require each country to rescore a sample of all other countries' assessments using its own national scoring rubrics. While desirable, this approach was judged to be impractical since it would have required the recruitment of scorers familiar with all the languages included in the assessment. As a compromise, countries agreed to exchange at least 300 randomly selected booklets with another country sharing the same test language.

In three cases, such an exchange was not possible. The presence of a multi-lingual Pole on the Canadian team enabled Statistics Canada to conduct the rescore of the Polish sub-sample. In Sweden and the Netherlands, however, no such opportunity existed. Therefore, in both of these countries, a second team of scorers was recruited and trained to independently rescore the requisite sub-sample using the Canadian scoring rubrics.

In all cases where serious discrepancies were identified, countries were required to rescore entire items or discrepant code pairs. The following sections provide detailed information on the final intra- and inter-country scoring reliabilities achieved for all items, including those that were dropped from the international scaling or were allowed to have unique item parameters.

Intra-country rescoring

A variable sampling ratio procedure was set up to monitor scoring accuracy. At the beginning of scoring, almost all responses were rescored to identify inaccurate scorers and to detect unique or difficult responses that were not covered in the scoring manual. After a satisfactory level of accuracy was achieved, the rescoring ratio was dropped to a maintenance level to monitor the accuracy of all scorers. Average agreements were calculated across all items. To ensure that the first and second scores were truly independent, certain precautions had to be taken. For example, scorers had to be different persons, and the second scorer could not be able to see the scores given by the first scorer.

Scorers who received identical training within a country are expected to be more consistent with one other than with scorers in other countries. This expectation was confirmed, as shown in Table 5.3. Most of the rescoring reliabilities were above 97 percent. It is important to note that the results were well within the statistical tolerances set for the IALS study and considerably better than those realized in other large-scale studies using open-ended items.

Table 5.3: **Intra-country rescore reliabilities, by country/language group**

Country	No. of booklets rescored	Average percent agreement
Canada/English	688	97
Canada/French	313	97
France	504	99
Germany	268	94
Netherlands	599	97
Poland	589	99
Sweden	301	98
Switzerland/French	1,187	98
Switzerland/German	1,143	97
United States	not applicable	not applicable

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Since intra-country rescoring was used as a tool to improve data quality, score updates were not made to the database. In other words, the agreement data presented here indicate the minimum agreement achieved in scoring. After intra-country reliabilities were calculated, a few scorers were found to be unreliable. These scorers either received additional training or were released. Where scores and rescores differed, the first scores were replaced with correct scores if the inaccuracy was due to a systematic error on the part of the first scorer. In some cases, the scoring guide was found to be ambiguous. In such cases, the scoring guide was revised and the first scores were changed to reflect the revisions, but the second scores were not altered. The second scores were never replaced, even if they were subsequently found to be erroneous.

In sum, the first scores reflect changes and corrections resulting from lessons learned in the intra-country rescoring analysis. The first scores are therefore more accurate and consistent than the second scores, which retain errors and thereby underestimate the rescore reliabilities somewhat. The extent to which the reliabilities are underestimated must be very small, however, given that most of the reliabilities are above 97 percent. These values indicate that very consistent scoring was achieved by all the participating countries.

Inter-country rescoring

Even after ensuring that all scorers were scoring consistently, fixing ambiguities in the scoring guides, and correcting any systematic scoring errors, it was still necessary to examine the comparability of scores across countries. Accurate and consistent scoring within a country does not necessarily imply that all countries are applying the scoring guides in the same manner. Scoring bias may be introduced if one country scores a certain response differently from the other countries. The inter-country rescorings described in this section were undertaken to ensure scoring comparability across countries.

As noted earlier, responses to the IALS assessment items were scored by each country separately. To determine inter-country scoring reliabilities for each item, the responses of a subset of examinees were scored by two separate groups. Usually, these scoring groups were from different countries. For example, a sample of test booklets was scored by two groups who scored Canada/English booklets and United States booklets. Inter-country score reliabilities were calculated by Statistics Canada, then evaluated by ETS. Based on the evaluation, every country was required to introduce a few minor changes in scoring procedures. In some cases, ambiguous instructions in the scoring manual were found to be causing erroneous interpretations and therefore lower reliabilities.

Using the inter-country score reliabilities, researchers can identify poorly constructed items, ambiguous scoring criteria, erroneous translations of items or scoring criteria, erroneous printing of items or scoring criteria, scorer inaccuracies, and, most important, situations in which one country consistently scored differently from another. In the latter circumstance, scorers in one country may consistently rate a certain response as being correct while those in another country score the same response as incorrect. This type of score asymmetry must be eliminated before the IRT scaling is performed. ETS and Statistics Canada identified such items, while the country in which the scoring problem occurred investigated the plausible causes for such systematic bias in scores. Where a systematic error was identified in a particular country, the original scores for that item were corrected for the entire sample.

Table 5.4 summarizes the inter-country rescore reliabilities, before corrections.

Table 5.4: **Inter-country rescore reliabilities, by country/language group**

Original country	No. of booklets rescored	Average percent agreement	Asymmetric items	Rescored by
Canada/English	158	97	1	United States
Canada/French	142	97	7	France
France	337	97	1	Canada/French
Germany	270	94	6	Switzerland/German
Netherlands	300	96	2	Netherlands*
Poland	300	97	2	Canada
Sweden	300	97	1	Sweden*
Switzerland/French	154	96	11	France
Switzerland/German	153	96	4	Germany
United States	315	97	0	Canada/English

* The Netherlands and Sweden carried out both inter- and intra-country rescoring internally due to a lack of available language experts in Dutch and Swedish. Separate groups were established to perform the rescoring.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Scoring asymmetries were defined as situations in which original scores were consistently different from rescores (e.g., a certain response was scored as correct in the original scoring but was consistently scored as wrong in the rescoring). Differences above 3 percent were used to identify asymmetric items. Six asymmetric items, erring in the same direction, were common to two countries: Canada/French and Switzerland/French, both of which had their responses rescored by France. It is not clear to why the asymmetry was not replicated when the French data were rescored by the Canada/French scorers. This type of consistent asymmetry on a set of items by French scorers might have been caused by a single scorer.

Because the purpose of the inter-country rescoring was to identify scoring difficulties and to determine whether the scoring standard was equivalent across countries, the second scores were not used to update the inter-country rescoring database. Accordingly, the data presented here reflect minimum agreement in the scoring procedures. The number of asymmetric items was calculated before corrections were made. Since the second scores were never changed, an enumeration of asymmetric items using the corrected first scores was not performed. The rates of agreement, most of which are within 1 percent, are nearly identical to those achieved in the intra-country rescoring. It is clear from this table that even the minimum agreements are very high.

5.5 Data Capture, Data Processing, and Coding

As a condition of their participation in the IALS, countries were required to capture and process their files using procedures that ensured logical consistency and acceptable levels of data capture error.

Specifically, countries were advised to conduct complete verification of the captured scores (i.e., enter each record twice) in order to minimize error rates. Because the process of accurately capturing the test scores is essential to high data quality, 100 percent keystroke validation was needed.

Each country was also responsible for coding industry, occupation, and education using standard international coding schemes (International Standard Industrial Classification, or ISIC; International Standard Occupational Classification, or ISOC; and International Standard Classification of Education, or ISCED). Further, coding schemes were provided for open-ended items, and countries were given specific instructions about the coding of such items so that coding error could be contained to acceptable levels. An excerpt from the Major Field of Study (MFS) coding scheme is attached as Appendix H.

To create a workable comparative analysis, each IALS country was required to map its national dataset into a highly structured, standardized record layout. In addition to specifying the position, format, and length of each field, this International Record Layout included a description of each variable and indicated the categories and codes to be provided for that variable. Upon receiving a country's file, Statistics Canada performed a series of range checks to ensure compliance to the prescribed format. When anomalies were detected, countries corrected the problems and submitted new files. Statistics Canada did not, however, perform any logic or flow edits, as it was assumed that participating countries performed this step themselves. An excerpt from the international record layout is provided as Appendix F.

5.6 Conclusions

The IALS study managers were well informed about IALS and its objectives, having been provided with background material and supported by information obtained at meetings and through discussions with Statistics Canada. With this knowledge, study managers made informed decisions about how best to administer the survey in their respective countries.

While most countries closely followed the data collection guidelines provided, some did deviate from the instructions. Known deviations from prescribed practice are described below.

First, two countries (Sweden and Germany) offered participation incentives to individuals sampled for their survey. The incentive paid in these two countries was trivial, however, and it is unlikely that this practice distorted the data. Research conducted in the United States indicates that small incentives have little impact on survey results. (See Chapter 6 for a discussion of incentives and motivation.)

Second, the doorstep introduction provided to respondents differed somewhat from country to country. Three countries (Germany, Switzerland, and Poland) presented the literacy test booklets as a review of the quality of published documents, rather than as an assessment of the respondent's literacy skills. A review of these practices suggests that they were intended to reduce response bias and were warranted by cultural differences in respondents' attitudes toward being tested. Given that the national study managers were well-briefed in the aims and goals of IALS, and were aware that the non-response rate should be kept to a minimum, we trust that they made informed decisions about how to introduce IALS to citizens of their countries.

Third, there were differences across the countries in the way in which interviewers were paid. No guidelines were provided on this subject, and the study teams therefore decided what would work best in their respective countries. It should be noted that the difference in payments across the countries does not take into account the purchasing power of the money in each country.

Fourth, several countries adopted field procedures which undermined the objective of obtaining completed background questionnaires for an overwhelming majority of selected respondents. Two examples are illustrative. In Germany, contract terms with the data collection agency induced interviewers to generate higher proportions of total non-respondents and of respondents providing cognitive data. While it is difficult to estimate the net effect of such a deviation, the evidence suggests it is negligible. In the Netherlands, limited provisions were made to provide language interpretation in sampled households containing non-Dutch speakers. Non-Dutch speakers do not appear to be underrepresented in the Dutch sample, however. We therefore conclude that the data collection agency found ways to cope with this restriction.

As a result of differences in data collection procedures, there are significant differences across the countries in the proportion of the total samples which are subject to imputation. It was therefore recommended that the imputation of literacy proficiency for non-respondents employed in the NALS study be retained in order to reduce bias in national estimates in those countries with cases to adjust. The overall impact of the adjustment is small, affecting the rank order of only two countries: the United States and Germany. Additional indirect evidence of the validity and comparability of the national datasets is provided by the principal component analysis which underlies the imputation procedure used in the IALS. Specifically, the Eigen values exhibit stable relationships and patterns across countries which outweigh the response differences observed.

The IALS countries all appear to have met the challenge of scoring the assessments reliably. Nonetheless, it might be useful to rescore samples of Swedish, Dutch, and Swiss booklets to eliminate the possibility of scoring error entirely. The assessment's psychometric performance and the quality of the national data from the international rescores suggest that the impact of scoring error was minimal. Because the impact of any given item is small, undetected scoring error would have to occur over many or all items to have any impact. The IALS countries' scoring results are exemplary compared to those of other international comparative studies and are well within the tolerances set by ETS prior to the launching of the assessment.

Capture, coding, and editing are standard procedures within any survey process. Most countries followed the IALS guidelines, verifying 100 percent of their data capture operation. The only two countries that did not comply with this recommendation conducted sample verifications, one country at 20 percent and the other at 10 percent. Given that capture errors are typically quite small, the error introduced at this stage should be negligible.

Each country coded and edited its own data. Armed with the detailed International Record Layout as well as the coding schemes for the ISIC, ISOC, MFS, and ISCED variables, countries were well prepared to code their files. Because in-depth data analyses for the initial IALS report have not revealed any startling anomalies in the data, any processing errors are probably randomly distributed and should not contribute significantly to errors in the data.

Three steps would help to control error in future rounds of the IALS. First, national study teams should be required to document key aspects of their study designs and submit these to Statistics Canada for review in advance of the pilot survey. Second, a detailed item adaptation guide should be developed to assist national study teams. Finally, an operations audit should be conducted just before the main study. This would enable study managers to identify changes to the survey that might compromise quality.

Chapter 6:

Incentives and the Motivation to Perform Well

Stan Jones

6.1 Introduction

Individuals who participated in the IALS knew that their scores would have no effect on their lives. In other words, it was a "low stakes" test for them. Participants probably had many different motivations for taking the test: some did so out of a sense of duty, others out of curiosity. If there were systematic differences in individual motivation among the countries, these differences might bias the results. This chapter explores the possible sources of such bias and examines data that allow us to assess the effectiveness of controls put in place to limit the impact of these sources of bias.

It is important to distinguish between two different types of motivation because they have different effects and because different types of evidence are examined in assessing their impact on the survey results. The first type is motivation to participate in the study. Some individuals contacted by interviewers were unwilling to participate in the IALS. Motivation of this kind can affect the survey coverage and influence the representativeness of the study population. In some respects, this source of potential bias is similar to flaws in sample design, and should be assessed in the same way. Section 6.2 looks at incentives and motivation to participate.

A second type of motivation is the motivation to make one's best effort in answering the survey tasks. Because there were no negative consequences for individuals if they performed poorly in the IALS, some may not have tried as hard as others. Thus, we might question whether the scores truly represent participants' abilities, or whether participants might have scored better (or worse) under different circumstances. Within any particular country, individuals undoubtedly differed in motivation to do well. What we wish to know is whether this type of motivation differed systematically from country to country because this, too, could be a source of bias. Section 6.3 discusses motivation to do well in the assessment.

6.2 Participation Bias

As discussed in Chapter 5, two IALS countries offered explicit incentives to encourage sampled individuals to take part in the survey and thereby increase survey response rates. Germany gave each participant a ticket to a lottery which offered a modest reward to the winner and lesser prizes to others. Sweden provided a nominal reward to participants. No other country offered monetary incentives to participants.

We have no direct evidence of the effectiveness of the incentives in these two countries because we have no data indicating what the response rates would have been if no incentives were provided. As part of the U.S. National Adult Literacy Survey (NALS), however, researchers in the United States conducted a large-scale experiment on the effect of incentives on survey participation (Mohadjer, Berlin, Reiger, Waksberg, Rock, Yamamoto, Kirsch, and Kolstad 1994). In the NALS pilot study, some participants were offered incentives of $20USD, others were offered $35USD, and still others were given no incentive. The study found that:

- "[T]he incentives were most effective in improving response rates for people with low educational attainment and for minority populations who are frequently underrepresented in national household surveys."
- The main effect of increased participation of these groups was to provide a better estimate of the skills of individuals with low education. "If incentives had not been used, self-selection factors resulting from reduced response rates among persons with lower levels of educational attainment would have *overestimated the literacy levels in the United States* for both the total population and major subgroups." (Italics added.)
- " ... [T]here was no significant relationship between incentive level [$0, $20, $35] and proportion of items attempted ... Once an individual had agreed to participate, level of incentive did not seem to have an impact on the individual's motivation to perform as measured by the number of tasks attempted."

The United States study suggests that incentives do increase survey participation rates, particularly among less educated adults, and that the use of incentives may therefore yield more precise estimates of the abilities of these adults. The impact of participation incentives on overall estimates is likely to be minimal, however. If anything, country scores may be lowered slightly when incentives are used, due to higher participation rates among less literate persons. Because incentives to participate have no effect on the amount of effort that individuals devote to the test, they have little or no effect on individual scores.

We do not know whether the results of the United States incentive experiment described above are directly applicable to other countries. To investigate this question fully would require additional studies of the same complexity (and cost) as the one conducted in the United States. The following analyses suggest, however, that the experiences of IALS countries which used incentives were similar to the experience in the United States.

Table 6.1 compares the educational characteristics of the weighted and unweighted samples of the various IALS countries. In countries where the sample proportion of a given education group is smaller than the weighted proportion, there is an underrepresentation of that group in the sample. Conversely, in countries where the sample proportion is larger than the weighted proportion, there is an overrepresentation of that education group.

Table 6.1: Comparison of weighted and unweighted sample distributions, by country

Education	Canada*	France	Germany	Netherlands	Poland	Sweden	Switzerland	United States*
Less than secondary grad.								
Sample	45	29	59	44	63	33	13	28
Weighted	36	39	59	45	63	36	19	18
Secondary graduation								
Sample	29	43	25	32	22	43	60	37
Weighted	33	42	25	38	22	41	61	45
Postsecondary (non-univ.)								
Sample	16	10	4	-	7	13	13	11
Weighted	17	7	4	-	7	12	11	15
University								
Sample	10	19	12	24	7	11	14	24
Weighted	15	12	11	17	7	10	9	22

* Both Canada and the United States intentionally oversampled adults with lower levels of education.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

In Germany and Sweden, where incentives were used, the portions of the samples with low levels of education and with high levels of education are similar to those of the countries' populations as a whole. Conversely, in France and Switzerland, which offered no incentives, individuals with low levels of education are underrepresented in the samples, and individuals with high levels of education (university education) are overrepresented.

In the Netherlands and Poland, two countries without incentives, the sample proportion of adults with low levels of education is a close match for the general population. Adults with secondary education are underrepresented in the Netherlands sample. Because Canada and the United States chose to overrepresent adults with low levels of education, the sample-weighted distributions for these two countries cannot be compared with those for the other countries.

These data suggest that when incentives to participate in a survey are offered, they tend to lower the average scores for the country, not increase them. This is because incentives induce individuals with less education, and likely lower literacy, to participate. Although the evidence for the IALS study is indirect, it is consistent with the United States NALS findings.

In summary, it is unlikely that participation incentives biased the results in Sweden and Germany. Instead, incentives are likely to have yielded more precise estimates of the distribution of literacy skills in those countries.

6.3 Motivation to Perform Well

No IALS country offered participants incentives to perform well on the test. Nonetheless, there is a question as to whether participants in some countries were more motivated to do well than those in other countries. In this section we review some of the evidence from other studies on these kinds of incentives and then apply that experience to the available IALS data.

It is not surprising that individuals are more likely to do well on a test when they are offered rewards for achieving high scores than when they have no direct rewards. Evidence from the United States (Bishop 1995) suggest that students do, in fact, perform better when they are given external motivation to do so.

Three pieces of evidence can be used to determine whether there were motivational differences across the IALS countries. We know that:

- Less motivated individuals are more likely to skip items, especially difficult items. If individuals are indifferent as to whether they score well on the test, they are unlikely to make an effort to answer difficult items, choosing instead to simply skip them as they were told they could do.
- Less motivated individuals are more likely to quit the test before the end, for the same reason that they are more likely to skip items. Although interviewers were trained to encourage respondents to attempt as many items as possible, they were also instructed to allow respondents to quit before completing the test if they wished to do so.
- Less motivated individuals spend less time on each task. Those who do not care much whether they complete a task correctly are likely to give the first answer that comes to mind and to spend relatively little time checking the correctness of their answers. They are also less willing to work carefully through the more complicated tasks. As a result, such respondents spend less time on the whole test than do those who are careful to ensure that their answers are the best they can give.

The following sections describe analyses of each of these types of evidence using the IALS data. Although the results do not reveal any systematic motivational differences among participating countries, the data are not unequivocal. For example, less skilled individuals are likely to skip tasks and to quit early in the test, but they are also more likely to take their time because it is more difficult for them to determine the correct answers. A more detailed examination of the evidence and its relation to the validity of the survey results is provided in Chapters 11 and 13.

6.4 Omit Rate

As the results in Table 6.2 indicate, the omit rate—that is, the percentage of tasks not answered—varied widely across the IALS countries and across the literacy scales. The lowest omit rate observed, 1.39, occurred in the quantitative component of the Canadian-English survey. The highest, 10, occurred in the prose portion of Poland's survey. The average omit rates across the IALS countries were 5.02 for the prose tasks, 3.43 for the document tasks, and 3.80 for the quantitative tasks.

Table 6.2: **Omit rate, by literacy scale and by country/language group**

Scale	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States	Average
Prose	2.04	4.35	5.80	4.12	3.69	10.0	8.51	3.94	4.15	3.52	5.02
Document	1.49	4.21	4.17	1.95	2.64	7.79	5.32	2.40	2.65	1.73	3.43
Quantitative	1.39	4.06	4.59	2.52	3.05	7.80	6.89	2.15	3.02	2.48	3.80

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

There appears to be a moderate relationship between a country's omit rate and its average literacy scores. These correlations are -.37 on the prose scale, -.53 on the document scale, and -.40 on the quantitative scale.

If omitting items is an indication of motivation, however, then motivation must not be highly correlated with scores. The country with the highest average literacy scores, Sweden, also had the second highest omit rate, and this rate was considerably larger than that in France, which had the third highest rate. Further, the country/language group with the lowest omit rate, Canadian English, had literacy scores in the middle range.

Even within a country there is no consistent pattern. Canadians who took the test in French had a higher omit rate on all scales than did Canadians who took the test in English, reflecting a difference in the ability distribution. In Switzerland, however, those who took the test in German had a consistently higher omit rate than those taking the test in French, but there was no general difference in ability between these two populations.

In sum, although the omit rate differed across the IALS countries, there is no strong evidence that low motivation leads to a high rate of omitted items and, consequently, to low test scores.

6.5 Not Reached Rate

The effect of quitting before finishing the test is mitigated by the fact that respondents were not penalized for not answering items. Because the IALS employed IRT scoring, which considered only those items attempted or omitted in the middle of the test, failure to complete the test did not result in a lower score. It merely decreased the precision of the score. Therefore, if the only effect of lower motivation was that respondents did not complete the test, the overall effect of differences in motivation would primarily be differences in precision, not differences in scores. Nonetheless, differences in uncompleted items may be viewed as a general indicator of differences in motivation.

To assess the impact of respondents stopping before the end of the test, we must consider several natural stopping points in the test. Some respondents quit at the start of the test, immediately after answering the background questions. (We can think of these as "not started" cases.) In such cases, it is likely that the individuals agreed to participate in the survey without fully understanding what they would be asked to do, and once they discovered that taking a test would be required, they chose not to continue.

As Table 6.3 indicates, the highest "not started" rates were in Poland, the United States, the French-speaking group in Canada, and France. It would be inappropriate to interpret non-starting as a measure of motivation to perform well. Rather, this should be viewed as a form of refusal to participate, or a type of non-response.

Table 6.3: **Not reached rates, by country/language group**

	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States	Average
Not started rate	15	22	22	6	3	29	5	14	15	27	16
Not finished rate	4	8	10	9	7	15	10	7	17	7	9

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

It is noteworthy that the United States had a higher "not started" rate in IALS than in its national survey, NALS, where participation incentives were used. As previously noted, such incentives are most effective in getting sampled individuals to agree to begin the interview, but are not particularly effective at stimulating them to do well.

While the "not started" rate is an indication of willingness to respond to the test at all, motivation to do well on the test is indicated by the relative number of participants who begin the test but quit before finishing it. We can refer to this as the "not finished" rate. As shown in Table 6.3, Poland and the German-speaking group in Switzerland had the highest "not finished" rates. The next highest rates were in Sweden, France, and Germany. The connection between this measure of motivation and scores is weak, however: some of the countries with high "not finished" rates had high literacy scores and others had low literacy scores.

6.6 Time Taken on Test

Less motivated individuals tend to spend less time on the literacy tasks and tend to be less concerned about whether their answers are correct or not. As Table 6.4 shows, there were differences among the IALS countries in the average length of time taken to complete the test.

Table 6.4: **Average test duration in minutes, by country/language group**

	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States	Average
Average duration, in minutes	67.8	73.4	75.9	63.6	66.2	79.6	*	59.8	60.9	*	69.3

* Neither Sweden nor the United States collected information about test duration.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

While low motivation might result in short test times, those with high ability might also complete the test quickly. In any case, if time were evidence of motivation to do well, then motivation should have been high in France and Poland (where the average test durations were high) and low in Switzerland (where the average test duration was low).

It is more plausible that time is generally associated with ability. Specifically, countries with the lowest overall scores also recorded the longest average test times. If low motivation results in lower test times and lower scores, then the evidence from the IALS suggests that participants' motivation did not differ systematically from one country to the next. In particular, France and Poland recorded the longest test times, not the shortest—which is what we would expect if respondents in these countries were less motivated to do well than were respondents in other countries. Indeed, if the amount of time taken to complete the test is an indication of motivation, then the French and Polish respondents must have been highly motivated to do well.

6.7 Conclusions

The evidence suggests that there was no systematic difference across the IALS countries in motivation to perform well on the literacy test. The differences in omit rates, "not finished" rates, and test duration are more plausibly associated with ability than with motivation. If these indicators were associated with motivation, then motivation could not be connected with scores.

Some differences in motivation to participate were evident. Specifically, incentives increased the participation rate among lower educated adults, resulting in more precise estimates for low-scoring respondents and, in general, overall lower scores than would have been attained without the incentives. It should be noted, however, that the two countries that offered participation incentives achieved the highest average scores in the IALS study.

If individuals were offered specific incentives to do well on the test, their scores would undoubtedly be higher. For these incentives to bias the results, however, their impact on the results would have to differ from country to country, and the available evidence suggests that this does not occur.

It is worth considering what would be needed to show that incentives to perform well differ from country to country, and that specific incentives can differentially affect scores. Individuals from different ages and social strata would need to be included in the study, because the size of the incentive affecting score differences could vary from group to group. Also, individuals from a variety of countries would need to be included to test whether the incentives differ from country to country.

If individuals were tested twice, once with incentives and once without, the test design would have to control for learning. Specifically, half the subjects would take the test first with an incentive, then without an incentive. The other half would take the test first with no incentive, then with an incentive. A substantial time period would have to pass between the administration of the two tests. Such a study would require significant funds.

The evidence from IALS strongly suggests that participants' motivation to do well did not bias the study results. The data also suggest that incentives used to encourage participation tended to increase the response rate of adults with lower literacy skills, resulting in lower average scores than would be observed in a no-incentive survey.

Part III:

Construct Validity and Proficiency Estimation

Introduction to Part III

In presenting the results of large-scale surveys, it is important to articulate not only what has been measured (the domain or construct), but also how it has been measured and what conclusions may be drawn from the data. Considerable effort was devoted to ensuring that the IALS was a statistically reliable and unbiased assessment of adult literacy. As with other aspects of the study, numerous measures were implemented to meet this objective, as described in this final part of the report.

Item performance

A key assumption underlying the IALS is that it is possible to construct literacy test items which function in a psychometrically equivalent manner both within and across heterogeneous populations. This assumption relies on seminal work by Irwin Kirsch of Educational Testing Service (ETS) and Peter Mosenthal of Syracuse University, who have explored factors to explain the difficulty of various types of adult reading tasks. This research is summarized in Chapter 7. This chapter reveals that IALS is the product of a decade of intensive assessment development and refinement involving millions of dollars and more than 40,000 respondents. As a result of this earlier work, more was known about the validity and reliability of the IALS instruments before going into the field than has been the case in any previous international assessment.

The Kirsch/Mosenthal framework proposes three separate literacy domains:

- *prose literacy*, which reflects the ability to read and use texts of varying levels of difficulty which are presented in sentence and paragraph form;

- *document literacy*, which reflects the ability to use texts of varying difficulty which are presented as tables, charts, and graphs, etc., and
- *quantitative literacy*, which reflects the ability to perform arithmetic calculations of varying difficulty which are based on text materials.

In each literacy domain, the difficulty of adult literacy tasks is determined by the interaction of the attributes of the text with the task that is to be performed. Since the Kirsch-Mosenthal system captures more than 85 percent of the variance in difficulty on each of the literacy scales, it becomes a simple matter to predict the difficulty of any given literacy task and to devise a test to predict the ability of population subgroups with a high degree of precision.

Assessment tasks are selected to ensure that, among population subgroups, each literacy domain (prose, document, and quantitative) is well covered in terms of difficulty, stimuli type, and content domain. Such a design ensures that there is sufficient information to estimate literacy ability and item difficulty and avoids disadvantaging any group due to differences in familiarity with tasks.

The IALS item pool was developed collectively by participating countries. Items were subjected to a detailed expert analysis at ETS and vetted by participating countries to ensure that the items were culturally appropriate and broadly representative of the populations being tested. For each country, experts who were fluent in both English and the language of the test reviewed the items and identified ones that had been improperly adapted. Countries were required to correct problems detected during this review process.

The pilot survey

To ensure that all of the final survey items had a high probability of functioning well, and to familiarize participants with the unusual operational requirements involved in data collection, each country was required to conduct a pilot survey. Although the pilot surveys were small and typically were not based strictly on probability samples, the information they generated enabled ETS to reject items, to suggest modifications to a few items, and to choose good items for the final assessment. ETS's analysis of the pilot survey data and recommendations for final test design were presented to and approved by participating countries at a meeting in Hamburg, Germany.

Test design

To obtain the maximum amount of information in the limited amount of survey time available, IALS adopted what is known as a "balanced incomplete block" design for administering the cognitive items. According to this design, the item pool was divided into several blocks, each of which required approximately the same amount of response time. The items within each block covered all three literacy domains (prose, document, and quantitative) and reflected a range of difficulty and content. The blocks were compiled into booklets, three blocks to a booklet. Each block appeared in all possible positions in a booklet (first, second, and third) so as to ensure that no subpopulation would be unduly disadvantaged by the placement of a particular set of items.

Psychometric analysis of survey data

The remaining chapters in this report present data on different aspects of IALS psychometric performance.

Chapter 8 presents a provisional analysis of the discriminant validity of the three reading constructs underlying the IALS assessment. The results of this analysis indicate that the literacy scales measure independent yet highly correlated skill domains.

Chapter 9 explores the reasons why particular test items failed to perform well psychometrically. Although the number of such items is small, one would be concerned if they revealed an inherent weakness in the Kirsch/Mosenthal theoretical framework. The analysis reveals that cases of unique item parameters are directly attributable to inadequate scrutiny of the final test items, a straightforward matter that can be corrected in future survey rounds.

Chapter 10 discusses scaling procedures and describes the manner in which the provisional IALS scales were linked to the NALS scales, providing a common metric which permits reliable international comparisons of adult literacy. Analyses described in this chapter (e.g., chi square statistics, visual inspection of item characteristics curves) show that very few IALS items failed to exhibit good fit across the countries, indicating that an overwhelming majority of the items discriminated in a stable and powerful way over the surveyed populations.

Chapter 11 documents the statistical procedures used to construct reliable and comparable proficiency estimates using the IALS cognitive data. It also describes the procedures used to impute the literacy proficiencies of sampled persons for whom cognitive information is limited or missing.

Chapter 12 uses a series of predictive models to examine the plausibility of the IALS proficiency estimates, and Chapter 13 examines the issue of construct validity using a LISREL framework. The latter analysis, while far from complete, provides additional empirical support for the validity of the IALS constructs.

Chapter 7:

The Measurement of Adult Literacy

Irwin S. Kirsch, Ann Jungeblut, and Peter B. Mosenthal

7.1 Introduction

Historically, standardized objective tests provided a means of comparing the performance of one individual with that of a relevant group or groups. A robust theory of mental tests and measurement, including statistical theory and procedures, was developed over the years to facilitate appropriate test interpretation and use. A theory of construct validity and sophisticated theoretical networks specifying research design strategies have also evolved. All of this work has served us well in both individual assessment and large-scale surveys, enabling us to compare the performance of individuals and groups with that of normative groups and to examine individual and group differences.

In addition to this work focusing on *people* assessed, developments in criterion-referenced testing have yielded increased information about the *tasks* used to assess them. These developments have provided better tools to analyze task difficulty and improved our ability to predict future test performance. What has been missing to date is a way to focus on people and tasks together to enhance the interpretability and utility of test results as well as decisions and actions based on these results.

Item response theory (IRT) scaling provides a way to focus on both people and tasks together. An IRT scale requires a pool of tasks of varying difficulty which assess the same dimension. Recent advances in psychometric theory and procedures make it possible to establish a scale for a unidimensional pool of tasks in which the ordering of difficulty is essentially the same for everyone. The scaling procedures first place tasks along the scale on the basis of pretest results; then, on the basis of demonstrated performance, individuals can be assigned scores on the same scale.

BEST COPY AVAILABLE

Displaying both the performance of people and the difficulty of tasks along a common scale raises some interesting questions. Do tasks with similar score values share certain characteristics? To what extent do these characteristics, and the response consistencies of individuals, differ systematically from one end of the scale to the other? The systematic exploration of these issues yields important benefits, namely:

- increasing our understanding of variables that contribute to task difficulty,
- enhancing our ability to generate new tasks that more fully represent the domain(s) being assessed,
- establishing a context for defining the domain boundaries—that is, enhancing score meaning, and
- strengthening the links between testing, research, and practice.

Collectively, these benefits contribute to an improved theoretical framework that helps us to account for consistency in task responses, which frequently reflects the interaction among a set of identifiable variables. Rather than treating task responses as a conglomeration of specifics, these response consistencies are typically summarized in the form of scores or subscores. According to Messick (1989), as our concern ceases to center on discrete behaviors or isolated observations, and we concentrate on providing score meaning, we move toward the level of measurement (1989, p. 14).

This chapter traces the evolution of the theoretical framework used to construct large-scale literacy assessments, including the IALS, and to report on the results. Since 1985, Educational Testing Service (ETS) has conducted three large-scale literacy assessments for the U.S. Departments of Education and Labor, each of which has been based on the same definition of literacy. The following sections describe these studies in terms of their contributions to our expanding understanding of literacy and the practical application of this understanding in literacy measurement. The issues and empirical evidence presented address various aspects of validity.

7.2 The Young Adult Literacy Survey

In 1985, ETS conducted a survey for the U.S. Department of Education's Office of Educational Research and Improvement (OERI) to assess the literacy skills of 21- to 25-year-olds in the United States. The development of the young adult literacy survey (YALS) was led by an expert panel which adopted the following definition of literacy: *Using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential* (Kirsch and Jungeblut 1986).

Based on this definition, the framework used to develop assessment tasks for the young adult survey took a multidimensional approach to literacy. The assessment tasks covered three important and distinct areas: prose, document, and quantitative literacy. Further, the assessment tasks focused on a variety of printed *material* and a variety of *uses* of this material. *Material* refers to the linguistic form in which information is displayed, while *use* reflects why the reader engages in the task, or the type of information the person needs or is seeking. Theoretically, use influences both the strategies and cognitive operations engaged in completing a reading task.

The matrix in Figure 7.1 shows the interaction between the 12 materials and 5 uses reflected in the survey tasks. This interaction was expected to define task difficulty and thus the placement of tasks on the literacy scales. The three score scales were designed to range from 0 to 500, with a mean of 250 and a standard deviation of 50 points.

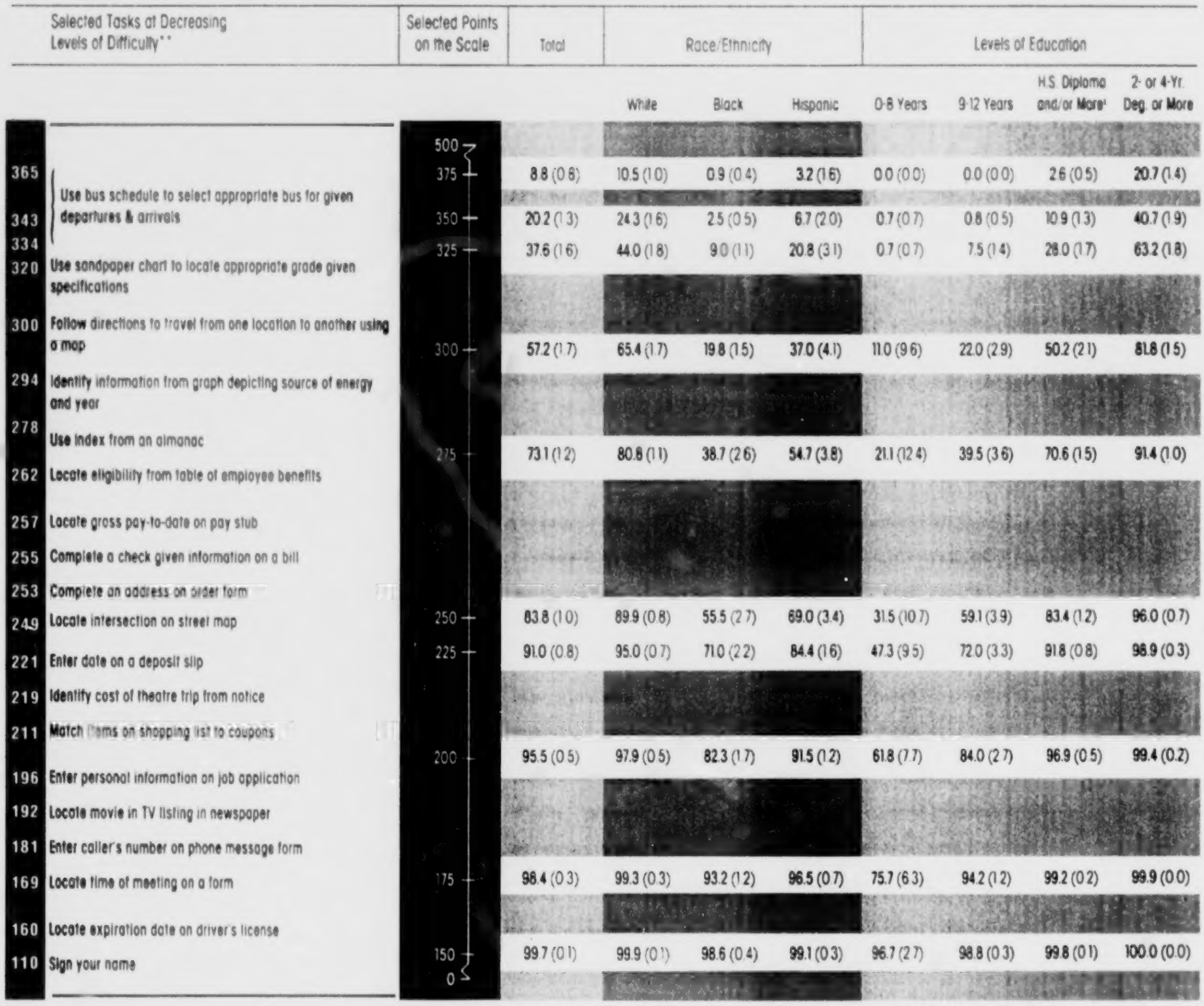
Figure 7.1: **Matrix of materials and uses for adult literacy tasks**

	Use				
Material	Knowledge	Evaluation	Specific information	Social interaction	Application
Sign/label			x		x
Directions					x
Memo/letter				x	
Form		x	x		x
Table		x	x	x	x
Graph		x	x		
Prose	x	x	x	x	
Index/reference			x		
Notice		x	x		x
Schematic			x		x
Ad			x		x
Bill			x		

Source: Educational Testing Service, Young Adult Literacy Survey, 1985.

Once the assessment tasks had been placed on the literacy scales, it was possible to summarize the survey results by identifying selected tasks at successive points on the scales and by indicating the percentages of adults (overall and in key subgroups) who scored at or above these points. This type of graphic presentation was expected to help give meaning to the scales and thereby increase the interpretability of the survey results. Figure 7.2 from the final report on the young adult survey (Kirsch and Jungeblut 1986, p. IV-18) shows the display for the document scale.

Figure 7.2: **Percentages of people and selected tasks at or above successive points on the document scale***



Selected Tasks of Decreasing Levels of Difficulty**	
365	
343	Use bus schedule to select appropriate bus for given departures & arrivals
334	
320	Use sandpaper chart to locate appropriate grade given specifications
300	Follow directions to travel from one location to another using a map
294	Identify information from graph depicting source of energy and year
278	Use index from an almanac
262	Locate eligibility from table of employee benefits
257	Locate gross pay-to-date on pay stub
255	Complete a check given information on a bill
253	Complete an address on order form
249	Locate intersection on street map
221	Enter date on a deposit slip
219	Identify cost of theatre trip from notice
211	Match items on shopping list to coupons
196	Enter personal information on job application
192	Locate movie in TV listing in newspaper
181	Enter caller's number on phone message form
169	Locate time of meeting on a form
160	Locate expiration date on driver's license
110	Sign your name

Selected Points on the Scale	Total	Race/Ethnicity			Levels of Education			
		White	Black	Hispanic	0-8 Years	9-12 Years	H.S. Diploma and/or More‡	2- or 4-Yr. Deg. or More
500								
375	8.8 (0.8)	10.5 (1.0)	0.9 (0.4)	3.2 (1.6)	0.0 (0.0)	0.0 (0.0)	2.6 (0.5)	20.7 (1.4)
350	20.2 (1.3)	24.3 (1.6)	2.5 (0.5)	6.7 (2.0)	0.7 (0.7)	0.8 (0.5)	10.9 (1.3)	40.7 (1.9)
325	37.6 (1.6)	44.0 (1.8)	9.0 (1.1)	20.8 (3.1)	0.7 (0.7)	7.5 (1.4)	28.0 (1.7)	63.2 (1.8)
300	57.2 (1.7)	65.4 (1.7)	19.8 (1.5)	37.0 (4.1)	11.0 (9.6)	22.0 (2.9)	50.2 (2.1)	81.8 (1.5)
275	73.1 (1.2)	80.8 (1.1)	38.7 (2.6)	54.7 (3.8)	21.1 (12.4)	39.5 (3.6)	70.6 (1.5)	91.4 (1.0)
250	83.8 (1.0)	89.9 (0.8)	55.5 (2.7)	69.0 (3.4)	31.5 (10.7)	59.1 (3.9)	83.4 (1.2)	96.0 (0.7)
225	91.0 (0.8)	95.0 (0.7)	71.0 (2.2)	84.4 (1.6)	47.3 (9.5)	72.0 (3.3)	91.8 (0.8)	98.9 (0.3)
200	95.5 (0.5)	97.9 (0.5)	82.3 (1.7)	91.5 (1.2)	61.8 (7.7)	84.0 (2.7)	96.9 (0.5)	99.4 (0.2)
175	98.4 (0.3)	99.3 (0.3)	93.2 (1.2)	96.5 (0.7)	75.7 (6.3)	94.2 (1.2)	99.2 (0.2)	99.9 (0.0)
150	99.7 (0.1)	99.9 (0.1)	98.6 (0.4)	99.1 (0.3)	96.7 (2.7)	98.8 (0.3)	99.8 (0.1)	100.0 (0.0)
0								

** Number indicating difficulty level designates that point on the scale at which individuals with that level of proficiency have an 80 percent probability of responding correctly.

* Numbers in parentheses are estimated standard errors.

‡ High school diploma and/or some postsecondary experience.

This figure shows, for example, that 90 percent or more of the total young adult population surveyed demonstrated proficiency on tasks at or below 225 on the document scale. These tasks, shown in the left-hand column, involved a variety of skills: entering the date on a bank deposit slip, identifying the cost of a particular theater trip from among those given in a notice, entering personal information on a job application form, locating the expiration date on a driver's license, or signing one's name. In addition to showing performance results for the total group, the figure indicates the results for various racial/ethnic groups and for individuals with various levels of educational attainment.

BEST COPY AVAILABLE

An anchoring procedure based on response probabilities was used to facilitate interpretation of the survey results and to attempt to identify factors associated with task difficulty. In this anchoring process, exemplar tasks were identified using IRT response probabilities. The criterion for selecting the exemplar tasks was that the response probability for individuals who scored around the same scale value as the task and the response probability for individuals who scored one standard deviation (50 points) lower had to differ by at least 30 points. The exemplars identified in this process represented a range of difficulty and also reflected a combination of skills thought to be associated with literacy proficiency.

On the prose scale, the skill or task characteristics identified in the anchoring process seemed to reflect three qualitatively different aspects of reading comprehension, as shown in Table 7.1. Each of these aspects was viewed as contributing to task difficulty, with significant overlap among the three. Interestingly, one aspect of prose comprehension—matching information in a question or directive with literal or corresponding (synonymous) information in a piece of text—also played an important role in defining task difficulty on the document scale and, to a lesser extent, the quantitative scale. Producing and interpreting text information also appeared to contribute to task difficulty on the prose scale.

Table 7.1: **Selected tasks and corresponding levels of difficulty defining the three aspects of the prose scale**

	Selected prose tasks		
Level of difficulty	Matching literal and corresponding information	Producing and interpreting text	Generating theme
500 —			
400 —			
375 —	-3-feature match from newspaper article (corresponding) **(386)***		-Generate theme from single unfamiliar metaphor **(376)***
350 —		-Interpret job-related benefit classification **(361)***	
325 —			-Generate theme from repetitive argument widely dispersed **(331)***
300 —			
275 —	-3-feature match from a page of text in an almanac (literal) **(273)***	-Interpret appliance warranty **(272)***	-Generate familiar theme from argument **(271)***
250 —			
225 —	-1-feature match from newspaper article (corresponding) **(205)***		
200 —		-Produce text with personalized background **(194)***	
0 —			

* Designates the point on the scale at which individuals with that level of proficiency have an 80 percent probability of responding correctly.

BEST COPY AVAILABLE

The most important characteristics associated with task difficulty on the document scale were the number of features that readers must identify in a question or directive and match to information in a document; the degree to which information in the question or directive is a literal or synonymous match with information in the document; and the number of exemplars or representations in the document having at least one feature in common with those in the question, thereby serving as distractors or possible correct answers. Table 7.2 provides information about task difficulty on the document scale (Kirsch and Jungeblut 1986, p. III-28).

Factors associated with task difficulty and performance on the quantitative scale appeared to be the type of arithmetic operation (addition, subtraction, multiplication, and division) required for a correct answer; the number or combination of operations needed; and the extent to which information specifying the operations is embedded in the textual material. Table 7.3 presents information about task difficulty on the quantitative scale (Kirsch and Jungeblut 1986, p. III-33).

Table 7.2: **Selected tasks and corresponding levels of difficulty defining the document scale**

Level of difficulty	**Selected document tasks**		
500 —			
400 —			
375 —			
	-6-feature match: Bus Schedule (Several Exemplars) **(346)***		
350 —			
	-4-feature match: Bus Schedule (Several Exemplars) **(316)***		
325 —			
300 —			
	-1-feature match: Filling-in Checks:	Dollars #	**(246)***
275 —	(Proced. Knowledge)	Date	**(244)***
		Dollar	**(242)***
		Pay to	**(241)***
250 —	-2-feature match: Pay Stub: Gross Year-to-date (Several Exemplars) Map: Location **(236)***		**(244)***
225 —			
200 —	-1-feature match: Pay Stub:	Current Net Pay	**(180)***
	(1 Exemplar) Meeting Room Form	Date	**(174)***
		Time	**(161)***
175 —			
150 —			
125 —	-1-feature match: Sign Name (Personal Knowledge - 1-Exemplar)	**(106)***	
0 —			

- Designates the point on the scale at which individuals with that level of proficiency have an 80 percent probability of responding correctly.

BEST COPY AVAILABLE

Table 7.3: **Selected tasks and corresponding levels of difficulty defining the quantitative scale**

Level of difficulty	Selected quantitative tasks
500 —	
	-2-Operations: Multiplication & Subtraction plus feature match **(479)***
400 —	
375 —	
	-1-Operation: Multiplication: Lunch Menu - Tip **(353)***
350 —	
	-2-Operations: Addition & Subtraction: Lunch Menu - Change **(335)***
325 —	
300 —	
	-1-Operation: Addition plus feature match involving entering and calculating checkbook balance
275 —	**(293, 290, 282, 282)***
250 —	
	-1-Operation: Addition using deposit slip **(236)***
225 —	
200 —	
0 —	

* Designates the point on the scale at which individuals with that level of proficiency have an 80 percent probability of responding correctly.

Understanding task difficulty

After the young adult literacy survey report was published, an ETS-funded study was undertaken to extend our understanding of the factors associated with task difficulty on the document scale. In this study (Kirsch and Mosenthal 1990), researchers applied a labor-, knowledge-, and time-intensive grammar to the 37 different documents (representing nine categories, such as tables, graphs, charts, and forms) as well as to the questions or directives given in the 61 document literacy tasks included in the young adult survey.

The grammar devised by Kirsch and Mosenthal (1990) relied on semantic-relation categories drawn from other propositional grammars. The process of applying the grammar to both the document stimulus materials and the associated task directives revealed interesting structural patterns and interactions. These patterns allowed Kirsch and Mosenthal to hypothesize about the variables underlying performance on the literacy tasks. These hypothesized variables related to the structure and complexity of the *document* or stimulus material, to the nature of the *task* (i.e., the structural relations between the document and the question or directive), and to the nature of the *processes* or strategies for relating information in the question or directive to information in the document. Within these three major categories of variables, initial analyses using percent correct statistics (rather than response probabilities) indicated 12 variables as significantly influencing the demonstrated document literacy scores of young adults.

Regression analyses further winnowed these 12 variables down to a set of five variables that proved to be significant (i.e., generalizable) across racial/ethnic groups and levels of educational attainment (Kirsch and Mosenthal 1990). Of these five significant variables, two were process variables: degrees of correspondence and type of information. Degree of correspondence refers to the level of difficulty involved in matching information in a question or directive to information in a document. Type of information refers to the level of difficulty involved in identifying the information provided in the document that is needed to answer the question or directive.

Two task variables—number of organizing categories and number of specifics required by the task—were also found to be significant. These variables represent the relation between the structure of the document and the information stated in the question or directive. For example, number of organizing categories consists of labels which serve to summarize or synthesize specific data or entries in a document. Similarly, the number of specifics relates to the number of entries or pieces of information the reader must process in order to perform the task correctly. The final significant variable, the number of specifics in the document, involves the length and complexity of the document itself.

One other process variable, plausibility of distractors, merits discussion here since it was significant for the total population. This variable refers to situations in which information in a text or document meets some but not all of the conditions set forth in the question or directive. Although this variable failed to attain significance for the various racial/ethnic and educational attainment groups, its profiles of zero-order correlations with percent correct scores were among the highest observed.

With the exception of young adults reporting zero to eight years of education, the variance accounted for by the five significant variables ranged from 89 percent for the total group assessed and for White young adults, to 81 percent for Black young adults as well as for those who reported dropping out of high school before earning a diploma. Only 56 percent of the variance in percent correct scores was accounted for in the group of individuals reporting zero to eight years of schooling. In general, then, the results yielded strong empirical validity evidence for both the evolving theory and for document score interpretation.

This study provided not only a theoretically based model of performance, but also a method for predicting task difficulty and identifying important cognitive characteristics of the young adult survey tasks. Although the results were not reported because of the small number of tasks, the approach just described was also applied to the prose and quantitative assessment tasks.

Understanding gained from these analyses of the young adult literacy survey data has made it possible to devise specifications for developing tasks targeted to specific areas of the literacy scales. For example, the coding scheme applied to the young adult survey data was useful in designing and developing new tasks for a subsequent survey of adult literacy, described below. Additional evidence of the validity of the theory and its applications to score interpretation would rest on the success of this task development work.

7.3 The U.S. Department of Labor Literacy Survey

A second large-scale literacy survey was conducted by ETS in 1989-90 for the U.S. Department of Labor (DOL). This survey assessed the literacy skills of Job Training Partnership Act and Employment Services/Unemployment Insurance program participants (Kirsch and Jungeblut 1992). It capitalized on the results of the young adult study as well as on an ongoing program of literacy research at ETS. The definition of literacy, the three literacy scales, and the expanded theoretical framework from the young adult survey all contributed to the work undertaken for the DOL assessment.

Although the earlier work provided an important starting point for the DOL survey, it was once again necessary to use the complex and labor-intensive grammar described earlier. Through a series of revisions and enhancements, researchers identified a set of variables that eliminated the need to use the grammar and greatly improved the utility of the coding procedures for others interested in literacy. Before discussing the DOL survey in depth, it will be useful to explain briefly how the current set of variables for prose, document, and quantitative tasks evolved from the original research.

As noted earlier, three process variables were identified through the initial young adult literacy survey research: degree of correspondence, type of information, and plausibility of distractors. In the work undertaken for the DOL survey, the plausibility of distractors variable was left unchanged. A new process variable, type of match, was constructed to merge the original type of information variable with the degree of correspondence variable. A third process variable, type of information, was added to indicate where information requested in a question or directive falls on a continuum from concrete to abstract.

Finally, it was necessary to respond to the concern that the variables identified and the associated variance accounted for in the young adult survey research might simply reflect the notion of "readability," which has a long history in theoretical and applied research. Accordingly, an estimate of readability was devised from the grammar for use with the document stimuli, and the Fry readability formula was used for the prose stimuli.

The following parts of this chapter evaluate the utility of the current framework as it relates to our ability to create new assessment tasks, increase our understanding of the variables contributing to task difficulty, and enhance score meaning.

Prose literacy

Prose literacy involves the knowledge and skills needed to understand and use information organized in sentence and paragraph formats. Given the range of text types organized in such formats, the DOL assessment used prose materials that were primarily expository (i.e., materials which described one or more states or actions) since such materials constitute much of the prose that adults read (Kirsch and Jungeblut 1986; Kirsch and Jungeblut 1992). Some narrative texts and poetry were also included. The prose materials were drawn from newspapers, magazines, books, brochures, and pamphlets, and were reprinted in their entirety, using the typography and layout of the original sources. As a result, the materials varied widely in length, density of information, and the use of structural or organizational aids, such as section or paragraph headings, italic or boldface type, and bullets.

Prose variables. In performing prose tasks, respondents must first identify "given" and "requested" information (Fisher 1981; Clark and Haviland 1977; Kirsch and Mosenthal 1992). *Given information* is information that is known and assumed to be true based on the way a question or directive is stated. *Requested information* in a question or directive is information being sought.

For example, consider the question, "In the past five years, how many times has Susan Butcher won the Iditarod Sled Dog Race?" In this instance, the given information is: "In the past five years, Susan Butcher won the Iditarod Sled Dog Race one or more times." The requested information is: "*How many times* did Susan Butcher win?" Prose tasks tend to be easy when the requested information is concrete, and become more difficult as the requested information becomes increasingly abstract. Hence, a task in which the requested information involves a person or thing (e.g., a "who" or "what" question) tends to be easier to process than a task that asks for a reason, purpose, or cause (e.g., "why"). We refer to these different degrees of abstractness as "type of information."

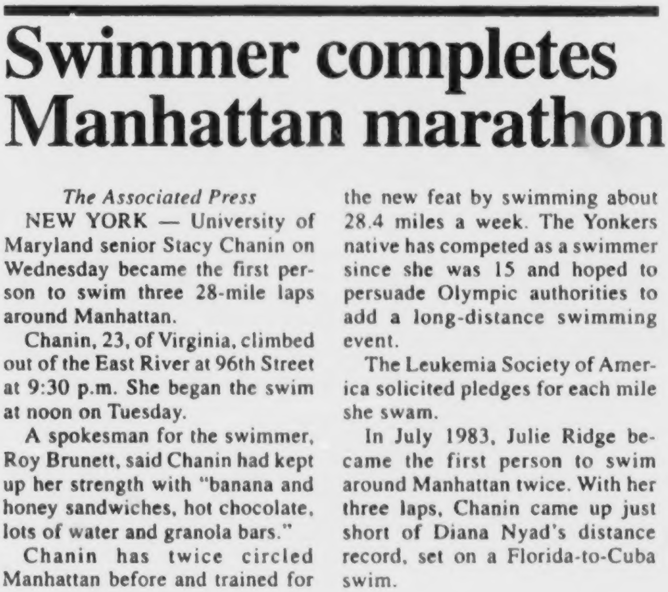
Another dimension of prose processing is matching information in a question or directive to corresponding information in a text. This involves the strategies of "locating," "cycling," "integrating," and "generating information." *Locating* tasks require the reader to find information in the text based on conditions or features specified in the question or directive. The match may be literal or synonymous, or the reader may need to make an inference in order to perform the task successfully. *Cycling* tasks require the reader not only to locate and match one or more features, but also to engage in a series of feature matches to satisfy conditions given in the question. *Integrating* tasks ask the reader to compare or contrast two or more pieces of information from the text. In some cases the information can be found in a single paragraph, while in others it appears in different paragraphs or sections. In *generating* tasks, readers must produce a written response by processing information from the text and by either making text-based inferences or drawing on their own background knowledge. These processes are represented by the variable "type of match."

A third dimension of prose processing pertains to situations in which the text contains "distracting information," or information which meets some but not all of the conditions specified in the question or directive. The more conditions that distracting information shares with an answer, and the more closely the distracting information is positioned relative to the correct answer, the more difficult processing becomes. We refer to this variable as "plausibility of distractors."

In addition to the three process variables, we included a fourth variable, Fry's (1977) readability, representing material complexity. We wanted not only to determine the extent to which this variable accounted for task difficulty, but also to provide a descriptor commonly found in the research literature on prose processing.

Each of the 44 prose tasks included in the DOL assessment was evaluated in terms of these four variables. A coding scheme was devised and applied to each of the prose literacy tasks. Coding for the variables "type of information" and "plausibility of distractors" ranged from 1 (easiest) to 5 (most difficult). Coding for the "type of match" variable was additive and ranged from 1 to a possible 20, although the DOL tasks ranged from 1 to 7.

Sample prose tasks. Using this coding scheme, let us examine several exemplar tasks which illustrate the range of task difficulty along the prose scale. One of the easiest prose tasks (with a scale value of 209) involved a short newspaper article about a marathon swimmer. This text reflects an eighth-grade Fry readability level. The directive asks the reader to "underline the sentence that tells what Ms. Chanin ate during the swim." To respond correctly, readers must recognize that the requested information is a "thing" (food). To identify this information, they must make a synonymous match between "ate" in the directive and "banana and honey sandwiches, hot chocolate, lots of water, and granola bars" in the text. Because there is no other mention of food in the text, there are no plausible distractors for the requested information. This prose task received a code of "1" for each of the three process variables.



Swimmer completes Manhattan marathon

The Associated Press

NEW YORK — University of Maryland senior Stacy Chanin on Wednesday became the first person to swim three 28-mile laps around Manhattan.

Chanin, 23, of Virginia, climbed out of the East River at 96th Street at 9:30 p.m. She began the swim at noon on Tuesday.

A spokesman for the swimmer, Roy Brunett, said Chanin had kept up her strength with "banana and honey sandwiches, hot chocolate, lots of water and granola bars."

Chanin has twice circled Manhattan before and trained for the new feat by swimming about 28.4 miles a week. The Yonkers native has competed as a swimmer since she was 15 and hoped to persuade Olympic authorities to add a long-distance swimming event.

The Leukemia Society of America solicited pledges for each mile she swam.

In July 1983, Julie Ridge became the first person to swim around Manhattan twice. With her three laps, Chanin came up just short of Diana Nyad's distance record, set on a Florida-to-Cuba swim.

A second task based on the same text has a scale value of 253. This task asks, "At what age did Chanin begin swimming competitively?" To answer, readers must recognize that the requested information is an amount (i.e., age) and then make a synonymous match to find the answer (15) in the sentence, "The Yonkers native has competed as a swimmer since she was 15...." What makes this task somewhat difficult is the fact that a distractor for the requested information appears earlier in the text; this distractor indicates that the swimmer's current age is 23. This task received a coding of 3 for type of match, 4 for plausibility of distractor, and 2 for type of information.

A third task (scale value of 247) was based on a long article on parenting by Dr. Spock. This article had an eighth-grade Fry readability level, but the text does not include any organizational aids. Readers are asked to identify one alternative to physical punishment. To respond, they can match the phrase "alternative to the use of physical punishment" to the phrase "other punishments parents can use." The text then lists a number of alternatives recommended by Dr. Spock. The type of match and plausibility of distractors variables each received a code of 2, and the type of information variable was coded a 3.

PARENTING

BY BENJAMIN SPOCK, M.D.

Have You Ever Wanted To Strike Your Child?

Don't do it! Dr. Spock believes that physical discipline can cause lasting resentment in a sensitive child and may make a naughty child a real behavior problem.

Almost all parents with whom I've ever discussed the issue of physical punishment acknowledge that they've had a strong impulse to spank their children at one time or another, whether they believed in doing it or not: for instance, when a small child breaks a valuable object she has been told not to touch, or when a somewhat older child of six or seven runs into the street and a car just misses hitting him, or when an eleven-year-old is caught stealing and then brazenly tries to lie her way out of it. And it's the rare parent who has *never* given in to the impulse to slap or spank.

Parents tend to punish their children the same way their own parents punished them — whether it's by spanking or scolding or reasoning or withholding privileges. In this way patterns of discipline — both good and bad — are passed from one generation to the next.

Why is it that physical punishment, whether used occasionally or frequently, is still widely accepted as a way of teaching children what is right and what is wrong? I think there are two reasons for this. The first is the belief that it is simply the correct way of handling certain kinds of misbehavior, such as those I've mentioned earlier. The second reason is even more powerful, and it has to do with the parent's *reaction* to the misbehavior: the wave of anger that sweeps over the parent when a child misbehaves, *especially* when there is an element of defiance in an act or in an attitude. The child's challenge to the parent's authority causes a spasm of panic: If the parent doesn't act quickly and with force, the child might get the upper hand and, as a result, the parent might lose some control permanently. While I don't believe that a child should be able to get away with such deliberate misbehavior, I do believe there are other effective ways a parent can discipline his or her child without resorting to physical punishment.

You may wonder why I feel that other forms of discipline are preferable to physical punishment. What convinced me that spanking isn't necessary was that, in years of pediatric practice, I discovered there were many families in which the children were never spanked — and yet these children were cooperative, polite and kind. In some of these families the parents had not been physically punished in childhood, either. In others, the parents remembered the humiliation of being hit or spanked and were reacting to a conviction that the spankings they had received as children had had the wrong effect.

The reaction of the parents who don't spank their children because they themselves were spanked is worth considering because it raises the question of whether physical punishment does any harm. It is obvious that, when applied occasionally by loving parents, it can't do *much* harm — after all, millions of good men and women have been brought up in this way. But I think there are better ways of influencing children. When physical punishment is used frequently, especially by irritable or harsh parents, its unfavorable effects are noticeably multiplied. I believe physical punishment teaches children that might makes right and helps to turn some of them into bullies. Physical punishment leaves some sensitive children with a lasting resentment toward their parents for having humiliated them in this way. It encourages other children to feel that violence is not really bad and to think of physical force as a way of solving problems or settling disputes. As adults we know it is not an effective way of solving problems or settling disputes.

To me the most important reason for trying not to use physical punishment is that, if it is effective, it makes the child behave out of fear of the pain and out of fear of your anger. I think it's preferable for children to do the right thing because they love their parents and want to please them — not because they fear them. Then, as the children grow up, go to school, get jobs, marry and raise a family, they'll carry over this same attitude of getting along well in life by loving people, wanting to please them and cooperate with them — and receiving that love and cooperation in return.

What about other punishments parents can use, such as taking away a beloved toy for a day or so? To me, the loss of a privilege seems better than the indignity of being hit.

Isolating a child who is out of control has been used effectively in good day-care centers. Sending a child to his room for a given period of time works just as well at home, but isolation should be used in a calm, friendly spirit, as a way of helping the child to cool off.

To me, the best way of ensuring good behavior is for parents to show children love and respect — from infancy — and to set a good example. Then children look up to their parents and want to please them.

When parents shout and hit, they thwart a child's natural desire to please her parents, because the child's love and respect for them has been diminished. In the long run, that makes the parents' job of disciplining their children all the more difficult.

You may think your children would never respond to anything as mild as a good example or a polite request. If they have been used to rougher forms of discipline, I'll admit that they will seem insensitive at first to gentler methods. But they will gradually come around. I've seen the transformation take place in a day-care center, where a thick-skinned misbehaver began cooperating with a gentle teacher after he slowly learned that he could trust her to be kind to him.

One approach you could use to get the attention of a child who has learned to ignore anything but the most extreme forms of correction would be to go to her immediately when she misbehaves, put your arm around her and say quietly, "When you do that, it makes me unhappy. Please don't do it again!" If misbehavior is consistently corrected in this fashion, not only will the child learn that she can't persist in whatever it is that she's doing wrong, but, more importantly, she will come to enjoy a better relationship with you and the impulse to misbehave will diminish. Of course, it takes a good deal of patience for a parent to make the shift to this kind of gentle discipline. But the results are well worth the effort.

Although Dr. Spock cannot answer readers' letters individually, he will respond to them in his column. Please address your questions to Department DW, Redbook, 224 West 57th Street, New York, NY 10019.

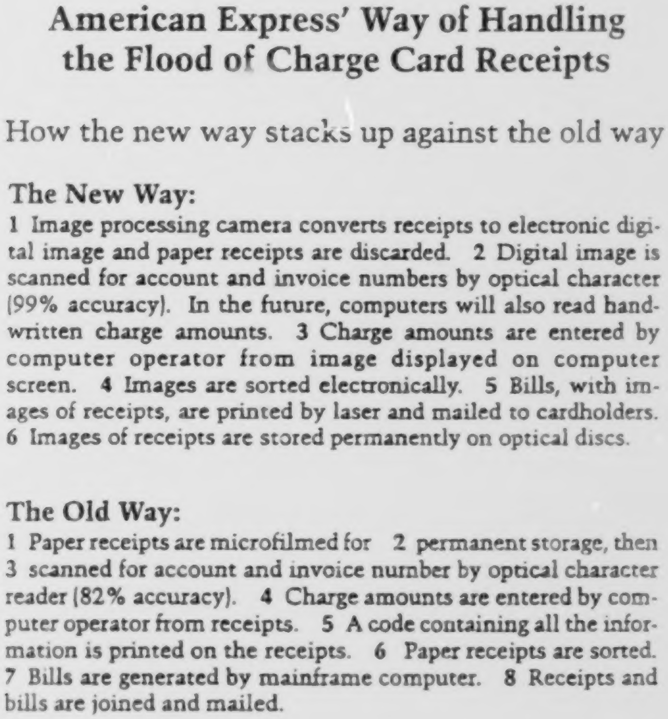
Benjamin Spock: "Have You Ever Wanted to Strike Your Child?" Reprinted from *Redbook* by permission of the publisher.

BEST COPY AVAILABLE

A more difficult prose task (scale value of 283) based on the same text requires the reader to "list the two reasons given by the author why physical punishment is still widely accepted as a way to teach children right and wrong." This task can be answered by locating the place in the text that begins, "I think there are two reasons for this. The first is .... The second reason is ...." For this task, the codes for type of match, plausibility of distractors, and type of information were 3, 2, and 4, respectively.

The most difficult task (scale value of 311) based on this text directs the reader to "Identify and list two reasons that Dr. Spock offers for not using physical punishment." While numerous statements throughout the article help satisfy the directive, much of the text deals with related concerns rather than direct summary statements. As a result, the reasons for not using physical punishment are embedded throughout the text and are not literally stated following a semantic cue such as "Two good reasons for not using physical punishment are ...." In addition, distracting information is more closely tied to words or phrases containing the necessary information for responding correctly. This task was assigned codes of 3 for type of match and plausibility of distractors and 4 for type of information.

A more difficult task (scale value of 346) directs readers to identify and list two similarities between the new and old ways in which American Express handles charge card receipts. Readers must compare and contrast several pieces of information to determine these similarities. This task was coded 4 for both type of match and type of information and 2 for plausibility of distractors. This piece of text also received an eighth-grade readability score based on the Fry formula.



American Express' Way of Handling the Flood of Charge Card Receipts

How the new way stacks up against the old way

The New Way:
1 Image processing camera converts receipts to electronic digital image and paper receipts are discarded. 2 Digital image is scanned for account and invoice numbers by optical character (99% accuracy). In the future, computers will also read handwritten charge amounts. 3 Charge amounts are entered by computer operator from image displayed on computer screen. 4 Images are sorted electronically. 5 Bills, with images of receipts, are printed by laser and mailed to cardholders. 6 Images of receipts are stored permanently on optical discs.

The Old Way:
1 Paper receipts are microfilmed for 2 permanent storage, then 3 scanned for account and invoice number by optical character reader (82% accuracy). 4 Charge amounts are entered by computer operator from receipts. 5 A code containing all the information is printed on the receipts. 6 Paper receipts are sorted. 7 Bills are generated by mainframe computer. 8 Receipts and bills are joined and mailed.

One of the most difficult prose tasks in the DOL survey requires the reader to identify two differences between the new and old ways of handling charge receipts. This task was coded 7 for type of match and 5 for both type of information and plausibility of distractors.

Validity evidence for the prose scale. One important piece of validation evidence can be obtained from analyses of the DOL survey tasks. As indicated earlier, the prose scale was not especially well defined in the young adult study; it contained only 15 tasks. Therefore, 29 new prose tasks were developed for the DOL survey. Each of the new tasks was coded using the revised theoretical framework described in this chapter. These variables were then used in regression analyses designed to predict the scale values of all of the tasks. Table 7.4 shows the results of these analyses for the new tasks as well as for the entire pool of DOL prose tasks.

Table 7.4: **Raw beta coefficients and standard errors for the predictive variables used in regression analysis of DOL prose tasks**

	New DOL prose tasks			All DOL prose tasks		
	Beta	Standard error	p^1	Beta	Standard error	p^2
Structure variables						
Readability	-.04	3.97	.93	2.14	2.35	.37
Process variables						
Type of match	12.08	5.94	.05	17.68	4.48	.00
Plausibility of distractors	28.53	5.79	.00	20.98	4.56	.00
Type of information	14.35	5.16	.01	12.19	4.28	.01
Total variance accounted for:						
R^2		81%			82%	
Adjusted R^2		78%			80%	

[1] df = 25
[2] df = 40

Source: Educational Testing Service, U.S. Department of Labor Literacy Survey, 1989-90.

Using the expanded theoretical framework to develop new tasks for the prose scale appears to have been very successful. The amount of variance accounted for in the new prose tasks (81 percent) as well as in the complete set of DOL prose tasks (82 percent) compares favorably with the results for document literacy tasks (81 to 89 percent) reported by Kirsch and Mosenthal (1990).

Document literacy

In addition to the ability to read and use prose materials, being literate in today's society means having the knowledge and skills needed to process documents, or information organized in matrix structures (i.e., in rows and columns). Documents include tables, signs, indexes, lists, coupons, schedules, charts, graphs, maps, and forms. In contrast to prose literacy, which tends to be the predominant form of literacy in schools, documents tend to be the principal form of literacy in non-school settings (Guthrie, Seifert, and Kirsch 1986). Documents serve many important functions in our daily lives (Bassett, Goodman, and Fosegan 1981; Burch and Grudnitski 1986). They enable us to perform important actions (e.g., applying for benefits, opening a charge account), make informed decisions (e.g., using a table of benefits to determine whether certain medical costs are covered), and record actions (completing a deposit slip or bill of sale, receiving a ticket for speeding).

Document variables. Document literacy tasks require readers to locate and match information in a question or directive to corresponding information in complex arrays, and to use this information in appropriate ways. Procedural knowledge may be needed to transfer information from one source or document to another, as is necessary in completing applications or order forms. As with the prose tasks, the matching involved in document tasks involves the strategies of locating, cycling, integrating, and generating information. These strategies are represented by the variable *type of match*. As with prose literacy tasks, success in processing documents also appears to depend on the ability to identify different *types of information* and the *plausibility of distractors*.

In addition to these three process variables, it was deemed important to provide an index of the "readability" of document stimuli. Since no such index is readily available for estimating document complexity, Kirsch and Mosenthal developed a method for estimating this complexity based on the grammar used in the earlier research.

The basic structural unit of documents is "simple lists" (Mosenthal and Kirsch 1989a). Such lists consist of a series of exemplars or items which belong to a common class of elements which in most instances are organized in terms of a more generic category or label. The documents used in the ETS literacy assessments reflect the ways in which a number of simple lists are organized to present more interdependent sets of information. These have been described as combined, intersecting, and nested lists (Kirsch and Mosenthal 1989a, 1989b; Mosenthal and Kirsch 1989b). The document readability variable ranges from 1 to 11. Included in this score is a number reflecting the type of document structure, the number of labels, and the numbers of items.

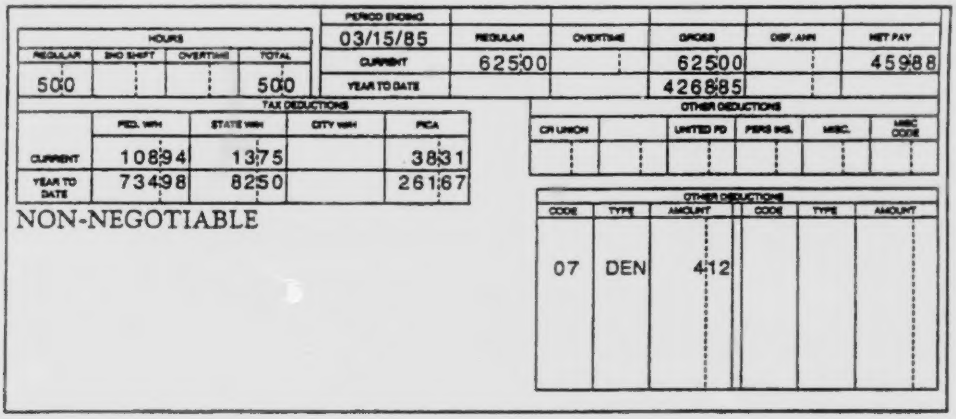
Sample document tasks. One of the easier tasks on the document scale (scale value of 198) requires the reader to look at a notice for a theater trip containing information about two plays. This document received a low "readability" score of 2. The reader is directed to circle the cost for a ticket and bus trip to see "On the Town." Although the reader must simply locate the line in the notice labeled "price" and circle the dollar amount associated with "On the Town," the cost given in the document for "Sleuth" serves as a plausible distractor. This task received a code of 1 for type of match, and a code of 2 for plausibility of distractors and type of information.

THEATER TRIP

A charter bus will leave from the bus stop (near the Conference Center) at 4 p.m., giving you plenty of time for dinner in New York. Return trip will start from West 45th Street directly following the plays. Both theaters are on West 45th Street. Allow about 1 1/2 hours for the return trip.

Time:	4 p.m., Saturday, November 20		
Price:	"On the Town"	Ticket and bus	$11
	"Sleuth"	Ticket and bus	$8.50
Limit:	Two tickets per person		

A more difficult task (275 on the document scale) directs readers to examine a wage and tax statement and select "gross pay for this year-to-date." If they fail to identify and match on both features, gross pay and year-to-date, then they are likely to respond with an incorrect amount based on distracting information. The readability of this document was coded 5. It was assigned a 2 for type of match and type of information and a 3 for plausibility of distractors.



HOURS			
REGULAR	2ND SHIFT	OVERTIME	TOTAL
5 00			5 00

PERIOD ENDING					
03/15/85	REGULAR	OVERTIME	GROSS	DEF. ANN	NET PAY
CURRENT	625 00		625 00		459 88
YEAR TO DATE			4268 85		

TAX DEDUCTIONS				
	FED. WH	STATE WH	CITY WH	FICA
CURRENT	108 94	13 75		38 31
YEAR TO DATE	734 98	82 50		261 67

NON-NEGOTIABLE

OTHER DEDUCTIONS					
CR UNION		UNITED FD	PERS INS.	MISC.	MISC CODE

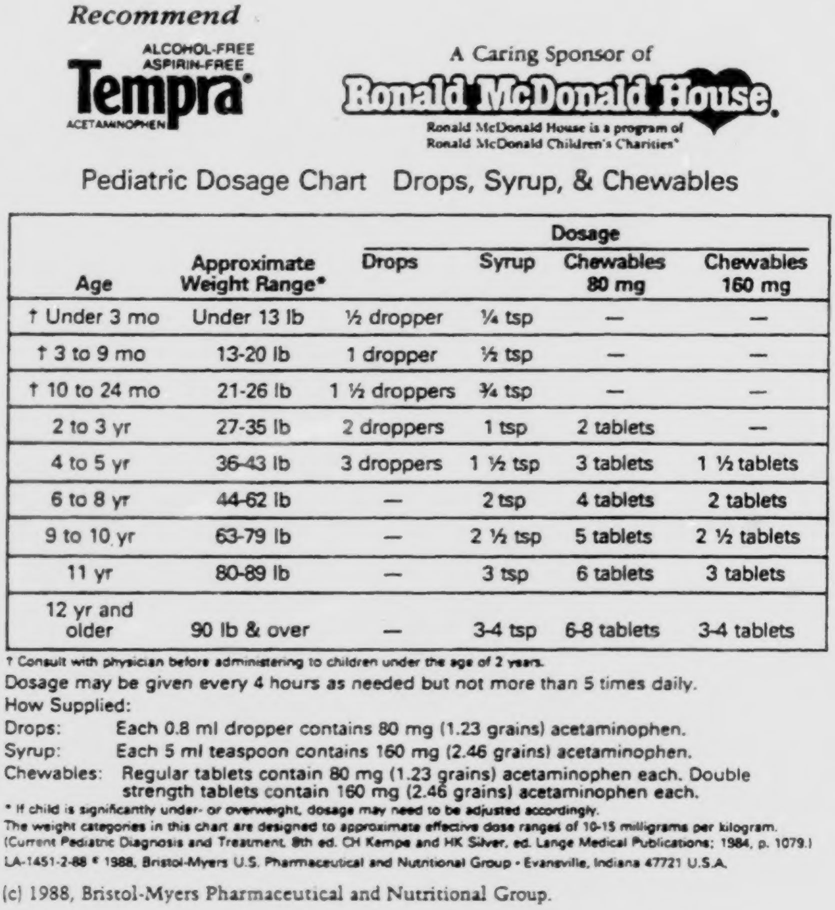
OTHER DEDUCTIONS					
CODE	TYPE	AMOUNT	CODE	TYPE	AMOUNT
07	DEN	4 12			

Reduced from original copy.

Another question based on this wage and tax statement was also expected to require a two-feature match—current and net pay—and therefore to have approximately the same scale value. This task was considerably easier (224), however. The codes assigned to the process variables indicate that little distracting information was present in the document. Each variable received a code value of 2.

Another document task (234) directs readers to look at a pediatric dosage chart and underline the sentence that indicates how often the medication may be administered. To respond successfully, readers must associate the word "administered" in the directive with the word "given" in the document by looking at information outside the table itself. The readability of this document was coded a 5. Although the type of match was assigned a code of 2, both plausibility of distractors and type of information were given codes of 3.

BEST COPY AVAILABLE



Recommend

ALCOHOL-FREE
ASPIRIN-FREE
Tempra®
ACETAMINOPHEN

A Caring Sponsor of
Ronald McDonald House.
Ronald McDonald House is a program of
Ronald McDonald Children's Charities®

Pediatric Dosage Chart Drops, Syrup, & Chewables

Age	Approximate Weight Range*	Dosage			
		Drops	Syrup	Chewables 80 mg	Chewables 160 mg
† Under 3 mo	Under 13 lb	½ dropper	¼ tsp	—	—
† 3 to 9 mo	13-20 lb	1 dropper	½ tsp	—	—
† 10 to 24 mo	21-26 lb	1 ½ droppers	¾ tsp	—	—
2 to 3 yr	27-35 lb	2 droppers	1 tsp	2 tablets	—
4 to 5 yr	36-43 lb	3 droppers	1 ½ tsp	3 tablets	1 ½ tablets
6 to 8 yr	44-62 lb	—	2 tsp	4 tablets	2 tablets
9 to 10 yr	63-79 lb	—	2 ½ tsp	5 tablets	2 ½ tablets
11 yr	80-89 lb	—	3 tsp	6 tablets	3 tablets
12 yr and older	90 lb & over	—	3-4 tsp	6-8 tablets	3-4 tablets

† Consult with physician before administering to children under the age of 2 years.

Dosage may be given every 4 hours as needed but not more than 5 times daily.

How Supplied:

Drops:	Each 0.8 ml dropper contains 80 mg (1.23 grains) acetaminophen.
Syrup:	Each 5 ml teaspoon contains 160 mg (2.46 grains) acetaminophen.
Chewables:	Regular tablets contain 80 mg (1.23 grains) acetaminophen each. Double strength tablets contain 160 mg (2.46 grains) acetaminophen each.

* If child is significantly under- or overweight, dosage may need to be adjusted accordingly.

The weight categories in this chart are designed to approximate effective dose ranges of 10-15 milligrams per kilogram. (Current Pediatric Diagnosis and Treatment, 8th ed. CH Kempe and HK Silver, ed. Lange Medical Publications; 1984, p. 1079.)

LA-1451-2-88 © 1988, Bristol-Myers U.S. Pharmaceutical and Nutritional Group • Evansville, Indiana 47721 U.S.A.

(c) 1988, Bristol-Myers Pharmaceutical and Nutritional Group.

A more difficult question (327) using this medicine dosage chart directs readers to determine how much syrup is recommended for a child who is ten years old and weighs 50 pounds. This task is difficult because readers cannot simply match literal or synonymous information to perform successfully since the weight given in the question is less than that of the typical ten year old according to the table. Instead, they must rely on prior knowledge or find the asterisked note relating to the column headed "Approximate Weight Range" to determine that the correct dosage is to be based on weight not age. If the reader approaches this task as a single literal match, then the age of the child is a highly plausible distractor and may lead to an incorrect response. The variable codes reflect this line of reasoning, with ratings of 4 for type of match and plausibility of distractors and a 2 for type of information.

Validity evidence for the document scale. As with the prose tasks, new document tasks were developed for the DOL survey to reflect various aspects of the theoretical framework that evolved from the young adult survey. Table 7.5 presents the results of regression analyses based on the new document tasks and the entire pool of document tasks for the DOL survey. Overall, 92 percent of the variance was accounted for in the new DOL document tasks and 87 percent in the set of all DOL document tasks. Given that variance accounted for in the young adult literacy study ranged from 81 to 89 percent, it appears that the integrity of the model was retained across the document tasks in the DOL survey.

Table 7.5: **Raw beta coefficients and standard errors for the predictive variables used in regression analysis of DOL document tasks**

	New DOL document tasks			All DOL document tasks		
	Beta	Standard error	p^1	Beta	Standard error	p^2
Structure Variables						
Readability	5.17	1.91	.01	1.39	1.10	.21
Process Variables						
Type of match	24.12	3.70	.00	24.46	2.28	.00
Plausibility of distractors	23.84	4.11	.00	22.71	2.44	.00
Type of information	-1.35	4.93	.79	9.09	3.15	.00
Total variance accounted for:						
R^2		92%			87%	
Adjusted R^2		91%			86%	

[1] df = 29
[2] df = 87

Source: Educational Testing Service, U.S. Department of Labor Literacy Survey, 1989-90.

Quantitative literacy

Since adults are often required to perform numerical operations in everyday life, the ability to perform quantitative tasks is an important area of adult literacy. To complete these types of tasks successfully, a respondent must perform arithmetic operations such as addition, subtraction, multiplication, or division either singly or in combination using numbers or quantities that are embedded in printed material.

At first glance, the skills involved in performing quantitative tasks might appear to be fundamentally different from those involved in processing prose and documents. An analysis of tasks along this scale shows, however, that processing printed information plays an important role in affecting the difficulty of quantitative tasks.

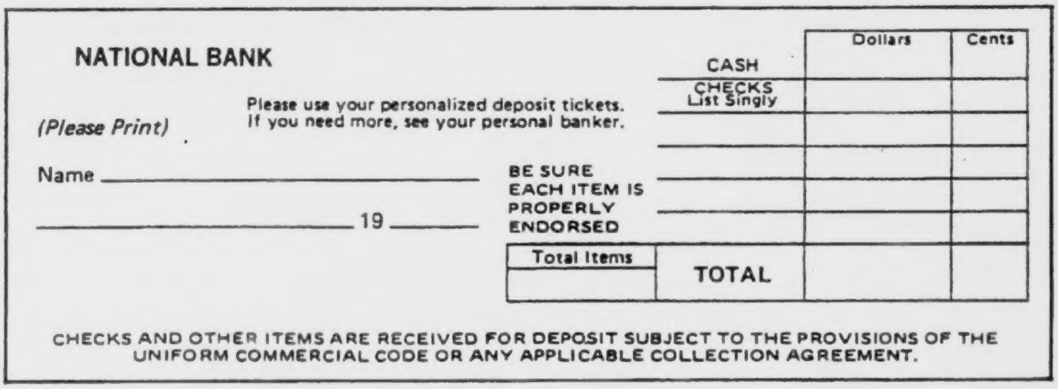
Quantitative variables. In general, it appears that many individuals can perform simple arithmetic operations using printed material when both the numbers and operations are made explicit. Yet, when the numbers for these same operations must be extracted from materials that contain similar but irrelevant information, or when the operations must be inferred, the tasks become increasingly difficult. To complete tasks on the quantitative scale, individuals are required to *match information* in a question or directive to information stated in one or more documents or pieces of text. In addition, quantitative literacy tasks may require respondents to deal with *plausible distractors* when extracting information for an arithmetic operation. Individuals are also required to process some type of information. While *type of information* varied for prose and document literacy tasks, requested information is almost always an amount in quantitative tasks. Most of the text materials for the quantitative tasks are documents of varying structural complexity.

Thus, the variables for the quantitative tasks are readability, type of match and plausibility of distractors—like those defined for the prose and document tasks. Two additional formulate variables are unique to this scale.

The first of these variables, *operation specificity*, refers to the process of identifying and sometimes entering numbers into an arithmetic expression, including determining the appropriate operation to be performed. Tasks tend to be more difficult when the numbers that must be identified appear in a document and are not in column format or adjacent to each other. Tasks also tend to become more difficult when the operation(s) is not specified and when the wording in the question or directive does not contain an explicit semantic relation statement such "how many" or "calculate the difference." The codes for the operation specificity variable ranged from 1 (easiest) to 9 (most difficult) based on a set of additive rules reflecting the various facets described here.

The second formulate variable, *type of calculation*, includes both the type of arithmetic operation (addition, subtraction, multiplication, or division) required and whether that operation must be performed alone or in combination. Tasks requiring two or more operations tend to be more difficult than those involving a single operation. Codes for this variable ranged from 1 (easiest) to 5 (most difficult).

Sample quantitative tasks. The least demanding quantitative task in the DOL survey (226) required the reader to enter and total two numbers on a bank deposit slip. In this task, both the number and operation could be easily identified, and readers simply had to add two decimal numbers that were already set up in column format. The numbers were stated in the directive, so the problem was framed for the reader. As a result, each of the process and formulate variables received a code of 1. The readability of the document was coded 2.



NATIONAL BANK

(Please Print)

Please use your personalized deposit tickets.
If you need more, see your personal banker.

Name ________________________

________________________ 19 ______

BE SURE
EACH ITEM IS
PROPERLY
ENDORSED

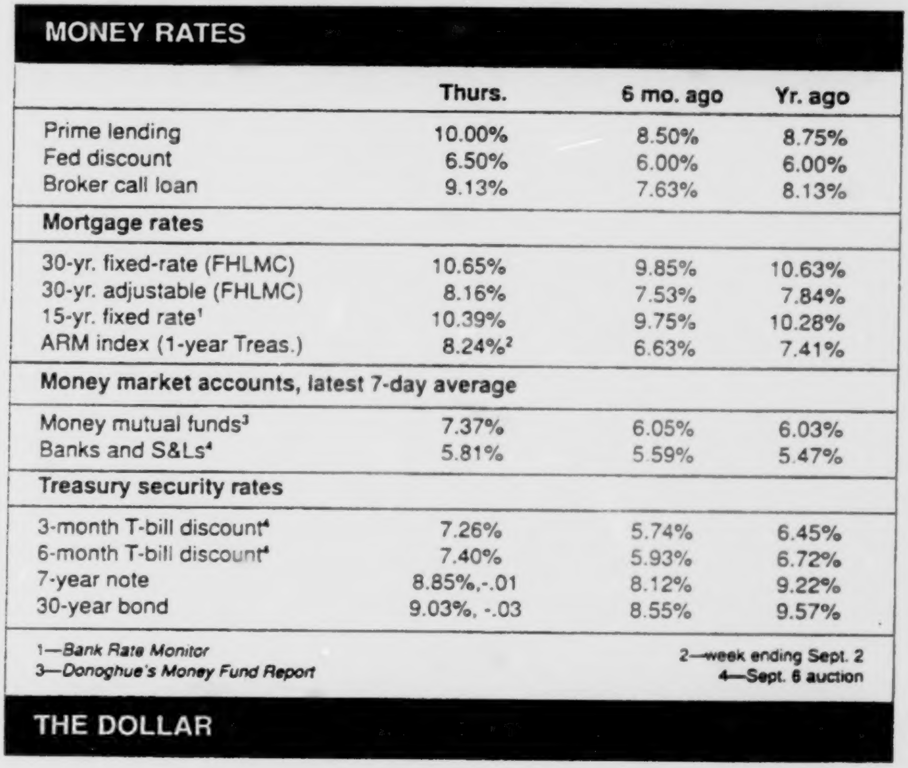
	Dollars	Cents
CASH		
CHECKS List Singly		
Total Items / TOTAL		

CHECKS AND OTHER ITEMS ARE RECEIVED FOR DEPOSIT SUBJECT TO THE PROVISIONS OF THE UNIFORM COMMERCIAL CODE OR ANY APPLICABLE COLLECTION AGREEMENT.

In similar tasks with higher scale values, the quantities may also have been easy to identify, but they were not explicitly given in the directive; instead, they had to be searched for and identified in the document. One such task (scale value of 265) required respondents to locate the appropriate shipping charges in a table, then enter the correct amount on an order form and calculate the total price for ordering office supplies.

Tasks at or about the 300 point on the quantitative scale still require a single arithmetic operation, but many of these tasks require respondents to identify two or more numbers from various places in the document. Typically, the numbers are not in column format, and the operation needed to complete the task is not explicitly stated in the directive or provided by the format of the document, as in the previous examples. Instead, the operation must be determined from arithmetic relation terms used in the directive, such as "how many" or "what is the difference."

One such task (283) requires the reader to look at a table of money rates to determine how much more interest would be earned in money market accounts provided by mutual funds than by S&Ls. This document received a readability score of 4, a code of 3 for type of match and plausibility of distractors, and a 2 for type of calculation and operation specificity.



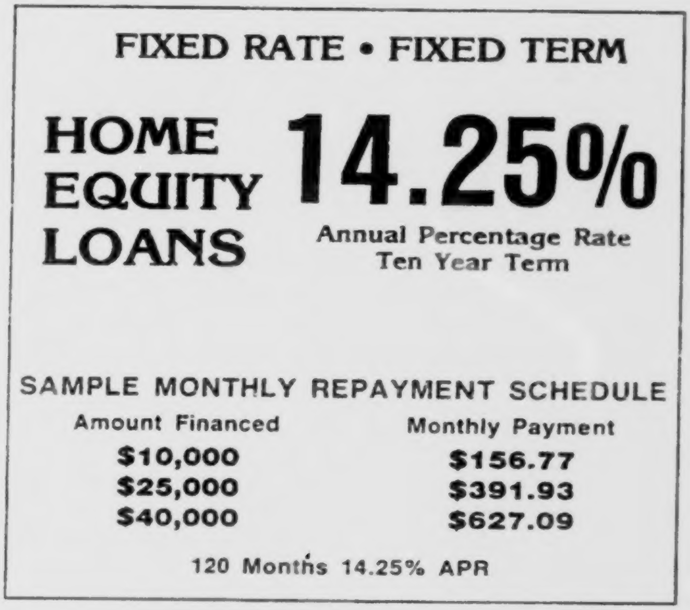
MONEY RATES

	Thurs.	6 mo. ago	Yr. ago
Prime lending	10.00%	8.50%	8.75%
Fed discount	6.50%	6.00%	6.00%
Broker call loan	9.13%	7.63%	8.13%
Mortgage rates			
30-yr. fixed-rate (FHLMC)	10.65%	9.85%	10.63%
30-yr. adjustable (FHLMC)	8.16%	7.53%	7.84%
15-yr. fixed rate[1]	10.39%	9.75%	10.28%
ARM index (1-year Treas.)	8.24%[2]	6.63%	7.41%
Money market accounts, latest 7-day average			
Money mutual funds[3]	7.37%	6.05%	6.03%
Banks and S&Ls[4]	5.81%	5.59%	5.47%
Treasury security rates			
3-month T-bill discount[4]	7.26%	5.74%	6.45%
6-month T-bill discount[4]	7.40%	5.93%	6.72%
7-year note	8.85%,-.01	8.12%	9.22%
30-year bond	9.03%, -.03	8.55%	9.57%

1—*Bank Rate Monitor*
3—*Donoghue's Money Fund Report*
2—week ending Sept. 2
4—Sept. 6 auction

THE DOLLAR

The quantitative tasks with the highest scale values (above 370) tended to require respondents to draw heavily on background information in order to identify both the quantities and operations needed to perform the task. The most difficult quantitative task (422) in this assessment, for example, required respondents to look at a newspaper advertisement for a home equity loan and then, using the information provided, explain how they would calculate the total amount of interest charges to be paid.



FIXED RATE • FIXED TERM

HOME EQUITY LOANS

14.25%

Annual Percentage Rate
Ten Year Term

SAMPLE MONTHLY REPAYMENT SCHEDULE

Amount Financed	Monthly Payment
$10,000	$156.77
$25,000	$391.93
$40,000	$627.09

120 Months 14.25% APR

Validity evidence for the quantitative scale. Only 15 quantitative tasks were used in the young adult literacy survey. Twenty-eight new tasks were developed for the DOL assessment using the theoretical framework described earlier in this chapter. As shown in Table 7.6, the combined set of readability, process, and formulate variables accounts for 84 percent of the variance in scale values for the 28 new DOL quantitative tasks and 83 percent of the variance for the entire set of 43 quantitative tasks administered in the DOL survey.

Table 7.6: **Raw beta coefficients and standard errors for the predictive variables used in regression analysis of DOL quantitative tasks**

	New DOL quantitative tasks			All DOL quantitative tasks		
	Beta	Standard error	p^1	Beta	Standard error	p^2
Structure variables						
Readability	4.81	1.60	.01	4.11	1.59	.01
Process variables						
Type of match	1.25	4.58	.79	.06	3.82	.99
Plausibility of distractors	20.44	3.52	.00	21.21	3.69	.00
Formulate variables						
Type of calculation	11.56	3.07	.00	10.76	2.25	.00
Operation specificity	8.23	2.76	.01	9.57	2.44	.00
Total variance accounted for:						
R^2		84%			83%	
Adjusted R^2		81%			81%	

[1] df = 22
[2] df = 37

Source: Educational Testing Service, U.S. Department of Labor Literacy Survey, 1989-90.

Establishing literacy levels

In reporting the results of the DOL assessment, the authors sought to clarify the graphic presentation of the survey data in order to avoid the information overload induced by the figures and tables in the young adult survey report (see Figure 7.2). A related objective was to provide a clear explanation of researchers' expanding insight into the meaning of the literacy scales.

Specifically, there was growing empirical evidence that an ordered set of information-processing skills and strategies is involved in accomplishing various kinds of prose, document, and quantitative literacy tasks. More difficult tasks tend to feature more varied and complex information-processing demands than easier tasks. This suggests that literacy is neither a single skill suited to all types of tasks nor an infinite number of skills each associated with a particular type of task.

In examining the distribution of DOL tasks across the literacy scales—or, more precisely, the process variable code values for these tasks—researchers noted that the values tended to "shift" at several points along the scales. These points occurred at *approximately* 50-point intervals, beginning at about the 225 scale point. While most of the tasks at the bottom end of the scales had code values of 1 for the three process variables, tasks with scores around 225 were more likely to have code values of 2. Among tasks with scores around 275, most of the codes were 2s and an increasing number were 3s. Among tasks with scale scores around 325, at least one of the three variables had a code value of 4. Code values of 4 predominated among tasks at or about 375, although occasional values of 5 or 6 were also found in this range. At approximately the 375 scale point, most of the code values ranged from 5 to 7.

Although there were some variations across the literacy scales in the points at which the coding shifts occurred, the patterns were remarkably consistent. Further, as we see later in the chapter, this system accounts for much (though not all) of the score variance associated with performance on the three literacy scales. Based on these findings, researchers defined five levels of literacy proficiency, with the following score ranges:

Level 1: 0 to 225
Level 2: 226 to 275
Level 3: 276 to 325
Level 4: 326 to 375
Level 5: 376 to 500

Once the literacy levels were tentatively defined, based on the apparent shifts in code values, researchers sought to establish criteria that would account for the placement of tasks within these literacy levels. These criteria are presented in Table 7.7. As the table shows, 78 percent of the prose tasks, 89 percent of the document tasks, and 79 percent of the quantitative tasks met the defined criteria.

Table 7.7: **Numerical criteria assigned to task variables to distinguish literacy levels, and percentages of tasks meeting the criteria, by literacy scale**

	Prose		Document		Quantitative	
	Criteria*	% Agree	Criteria*	% Agree	Criteria*	% Agree
Level 1	1, 1, ≤2	100	1, ≤2, ≤2	87	1, 1, ≤2	100
Level 2	2, 2, 2, or 3, ≤3, ≤3	73	2, 2, 2, or 3, ≤3, ≤2	97	≤3, ≤2, ≤2	50
Level 3	4, ≤3, ≤3	80	≥4, ≤3, ≤3	87	≤5, ≤3, ≤3	82
Level 4	4, ≥4, ≤4	73	4, ≥4, ≤3	75	≥4, ≤4, ≤4	77
Level 5	5, ≥5, ≤5	75	5, ≥5, ≤5	86	5, ≥5, ≤5	80
Overall agreement		**78**		**89**		**79**

* The numbers in this column refer to the codes for the process variables.

Source: Kirsch and Mosenthal.

The next step in evaluating the utility of the five literacy levels for reporting purposes was to run regression analyses using levels rather than individual task scale values as the dependent variable. As shown in Table 7.8, the process variables that are significant in predicting literacy levels are the same as those that are significant in predicting specific task scale values. Moreover, the models used to predict literacy levels account for roughly the same amount of variance as those used to predict task scale values: 80 percent on the prose scale, 88 percent on the document scale, and 78 percent on the quantitative scale. These findings are somewhat surprising given that there are no tasks at the very high or very low ends of the scales; such restrictions in range typically produce smaller correlations.

Table 7.8: **Raw beta coefficients and standard errors for the predictive variables used in regression analysis of DOL levels for the prose, document, and quantitative scales**

	Prose levels			Document levels			Quantitative levels		
Prose levels	Beta	Standard error	p^1	Beta	Standard error	p^1	Beta	Standard error	p^1
Structure variables									
Readability	.03	.04	.53	.00	.20	.93	.07	.37	.07
Process variables									
Type of match	.19	.08	.03	.46	.04	.00	-.05	.09	.58
Plausibility of distractors	.49	.09	.00	.43	.04	.00	.48	.09	.00
Type of information	.25	.08	.01	.21	.06	.00	—	—	—
Formulate variables									
Type of calculation	—	—	—	—	—	—	.14	.05	.01
Operation specificity	—	—	—	—	—	—	.20	.06	.00
Variance accounted for:									
R^2		80%			88%			78%	
Adjusted R^2		78%			87%			75%	

Source: Educational Testing Service, U.S. Department of Labor Literacy Survey, 1989-90.

These data show that the literacy levels account for significant amounts of variance in performance on literacy tasks. They also provide evidence of the substantive aspect of construct validity, or the appropriateness of the theoretical model predicted to underlie consistencies in performance on literacy tasks. In other words, increasing levels of literacy proficiency reflect increasing complexity of the construct of interest rather than increasing sources of construct-irrelevant difficulty.

Based on this evidence of the validity of the five literacy levels, it was decided that the levels would be used to report the results of the DOL survey (Kirsch, Jungeblut, and Campbell 1992). In addition to presenting the percentages of adults in the DOL programs (JTPA and ES/UI) who scored in the various literacy levels, the report authors defined what it means to perform in these levels, based on an analysis of the knowledge and skills required by the tasks in each level. Finally, drawing on IRT-based analyses, it was possible to estimate the likelihood that individuals in each literacy level would be able to perform the average task in that level and in the other levels. Unlike a traditional test score, which provides a single estimate of ability, these probability estimates offer a richer and more accurate reflection of the range of tasks that a person can be expected to perform successfully.

In the report presenting the DOL results, these three types of information—the literacy level definitions, percentage distributions, and probabilities of successful performance on tasks in the various levels—were provided in a single table, one for each literacy scale. Table 7.9 shows how results were presented for the document scale, for example. According to this table, 37 percent of JTPA participants scored in the level 2 range (between 226 and 275) on the document scale, and the average scale score value of tasks in this level was 248. A person scoring at or about 250 (that is, in the level 2 range) has an 81 percent probability of being able to perform a document task in this level correctly, and an even higher likelihood of success (94 percent) on the tasks in level 1. On a level 3 task, however, this same person's probability of success drops to 54 percent, and on level 4 and 5 tasks it is only about 25 percent.

In other words, adults are highly likely to be able to perform the tasks in their own literacy level successfully, but their likelihood of success declines as task difficulty increases.

Table 7.9: **Summary of NALS item performance by level and by assessment**

	Document literacy		Average probability at selected levels						
Level	Description of document tasks at each of five levels	Average RP80 at level	200	250	300	350	400		Total
Level 1 0-225	Tasks at this level are the least demanding. In general, they require the reader to either locate a piece of information based on a literal match or to enter information from personal knowledge.	194	80	94	98	100	100	DOL: JTPA DOL: ES/UI YALS	14.1 (2.0) 13.1 (1.6)
Level 2 226-275	Tasks at this level begin to become more varied. Some still require the reader to match a single piece of information; however, tasks occur where the match is based on low-level inferences. Tasks at this level also begin to require the reader to cycle through information or to integrate information.	248	47	81	95	99	100	DOL: JTPA DOL: ES/UI YALS	37.3 (1.3) 30.1 (1.2) 24.2 (1.2)
Level 3 276-325	Tasks at this level tend to require the reader to either integrate three pieces of information or to cycle through materials in rather complex tables or graphs in which distractor information is present.	300	30	54	79	93	97	DOL: JTPA DOL: ES/UI YALS	35.4 (1.5) 35.9 (1.0) 39.7 (1.2)
Level 4 326-375	Tasks at this level continue to demand more from the reader. Not only are multiple-feature matching, cycling and integration of information maintained, the degree of inferencing is increased. Cycling tasks often require the reader to make five or more responses with no designation of the correct number of responses. Conditional information is also present and must be taken into account.	351	11	26	53	79	83	DOL: JTPA DOL: ES/UI YALS	12.2 (1.8) 18.5 (1.7) 24.0 (1.1)
Level 5 376-500	Tasks at this level require the most from the reader. The reader must search through complex displays containing multiple distractors, make text-based inferences, or use specialized knowledge.	405	15	23	37	60	79	DOL: JTPA DOL: ES/UI YALS	1.1 (0.4) 2.4 (0.5) 4.1 (0.6)

* The numbers in parentheses are estimated standard errors.

Source: Educational Testing Service, National Adult Literacy Survey, 1993.

In summary, the authors of the DOL assessment report sought to make the survey results more useful to researchers, policymakers, and others by providing more detailed information than had been included in the young adult literacy survey report, and by improving the clarity of the data displays. As the following section shows, the subsequent National Adult Literacy Survey offered additional opportunities to examine the validity of the theoretical framework described here, as well as to improve on score interpretation and reporting.

7.4 The National Adult Literacy Survey

The National Adult Literacy Survey (NALS), conducted by ETS in 1992 for the U.S. Department of Education, assessed the literacy skills of a nationally representative sample of individuals aged 16 and above. Because an important goal was to link the results of the NALS to those from the young adult and DOL assessments, the committees which led the development of the NALS decided that this survey should be based on the same definition of literacy that had guided the earlier surveys. They also recommended that the NALS results be reported in terms of the five literacy levels that had been developed for the DOL assessment.

To further investigate the validity of the theoretical framework described earlier in this chapter, researchers carried out new regression analyses using data from the pool of NALS tasks. This pool consisted of 41 prose tasks (26 of which were newly developed for this survey), 81 document tasks (33 of which were new), and 42 quantitative tasks (27 new) (U.S. Department of Education, in progress). Table 7.10 presents the results of these analyses for the new NALS tasks, and Table 7.11 provides parallel results for the entire pool of NALS tasks.

Table 7.10: **Raw beta coefficients and standard errors for the predictive variables used in regression analyses of new NALS prose, document, and quantitative tasks and levels**

	Task			Level		
	Beta	Standard error	p^1	Beta	Standard error	p^1
Prose						
Structure variables						
Readability	-.07	1.8	.68	-.02	.03	.59
Process variables						
Type of match	29.65	3.62	.00	.57	.07	.00
Plausibility of distractors	18.41	4.06	.00	.29	.07	.00
Type of information	10.91	4.35	.02	.15	.08	.06
Formulate variables						
Type of calculation	-	-	-	-	-	-
Operation specificity	-	-	-	-	-	-
Total variance accounted for:						
R^2		89%			89%	
Adjusted R^2		87%			87%	
^{1} df =		24			24	
Document						
Structure variables						
Readability	2.83	2.58	.29	.07	.06	.24
Process variables						
Type of match	17.09	3.57	.00	.34	.08	.00
Plausibility of distractors	28.15	5.30	.00	.63	.12	.00
Type of information	9.84	5.07	.07	.24	.11	.04
Formulate variables						
Type of calculation	-	-	-	-	-	-
Operation specificity	-	-	-	-	-	-
Total variance accounted for:						
R^2		88%			87%	
Adjusted R^2		85%			85%	
^{1} df =		20			20	
Quantitative						
Structure variables						
Readability	.33	3.18	.92	-.07	.05	.22
Process variables						
Type of match	-2.57	5.16	.62	.05	.09	.57
Plausibility of distractors	12.25	5.16	.03	.28	.09	.00
Type of information	-	-	-	-	-	-
Formulate variables						
Type of calculation	20.75	4.62	.00	.31	.08	.00
Operation specificity	16.64	3.84	.00	.32	.07	.00
Total variance accounted for:						
R^2		84%			85%	
Adjusted R^2		81%			82%	
^{1} df =		22			22	

Table 7.11: **Raw beta coefficients and standard errors for the predictive variables used in regression analyses of all NALS prose, document, and quantitative tasks and levels**

	Task			Level		
	Beta	Standard error	p [1]	Beta	Standard error	p [1]
Prose						
Structure variables						
Readability	.18	1.67	.00	.01	.03	.65
Process variables						
Type of match	28.86	3.45	.00	.50	.06	.00
Plausibility of distractors	16.09	3.63	.00	.26	.06	.00
Type of information	8.84	4.17	.04	.15	.07	.05
Formulate variables						
Type of calculation	-	-	-	-	-	-
Operation specificity	-	-	-	-	-	-
Total variance accounted for:						
R^2		87%			88%	
Adjusted R^2		86%			86%	
[1] df =		36			38	
Document						
Structure variables						
Readability	.09	1.07	.42	.00	.02	.92
Process variables						
Type of match	20.33	2.27	.00	.42	.04	.00
Plausibility of distractors	22.19	2.74	.00	.43	.05	.00
Type of information	11.69	3.05	.00	.20	.05	.00
Formulate variables						
Type of calculation	-	-	-	-	-	-
Operation specificity	-	-	-	-	-	-
Total variance accounted for:						
R^2		84%			86%	
Adjusted R^2		83%			86%	
[1] df =		76			76	
Quantitative						
Structure variables						
Readability	.78	2.81	.78	.03	.05	.64
Process variables						
Type of match	.58	4.24	.87	.04	.08	.64
Plausibility of distractors	9.66	4.17	.03	.21	.79	.01
Type of information	-	-	-	-	-	-
Formulate variables						
Type of calculation	14.43	3.21	.00	.21	.06	.00
Operation specificity	18.39	2.94	.00	.35	.06	.00
Total variance accounted for:						
R^2		81%			79%	
Adjusted R^2		78%			76%	
[1] df =		37			37	

The results show that, on each of the literacy scales, the variables that contribute to the predictive models for the NALS data are the same as those for the DOL data. Further, the amount of variance accounted for is nearly the same for both individual tasks and levels. As shown in Table 7.10, the amounts of explained variance for the prose levels (based on the new NALS tasks) are 89 percent on the prose scale, 87 percent on the document scale, and 85 percent on the quantitative scale. The analyses based on the entire pool of NALS tasks reveal comparable results, as shown in Table 7.11.

The similarity of results between the NALS and DOL task pools confirmed the appropriateness of interpreting the NALS results using the five literacy levels established to present the DOL assessment findings. Thus, the initial report summarizing the NALS findings (Kirsch, Jungeblut, Jenkins, and Kolstad 1993) presented the percentages of adults nationwide, and of those in various subpopulations, who scored in each of the five literacy levels on the prose, document, and quantitative scales. Like the DOL report display presented earlier (see Table 7.9), the NALS overview report also briefly defined these levels, based on an analysis of the knowledge and skills represented by the tasks in each level. These definitions are shown in Figure 7.3.

In summary, analyses based on data from the NALS show that the literacy levels account for significant amounts of variance in performance on literacy tasks and offer additional evidence of the validity of the theoretical framework described earlier in this chapter. An ordered progression of information-processing skills and strategies appears to be involved in performing various types of prose, document, and quantitative literacy tasks, and the literacy levels offer a useful way to characterize these patterns.

Figure 7.3: Description of the prose, document, and quantitative literacy levels

	Prose	Document	Quantitative
Level 1 *0-225*	Most of the tasks in this level require the reader to read relatively short text to locate a single piece of information which is identical to or synonymous with the information given in the question or directive. If plausible but incorrect information is present in the text, it tends not to be located near the correct information.	Tasks in this level tend to require the reader either to locate a piece of information based on a literal match or to enter information from personal knowledge onto a document. Little, if any, distracting information is present.	Tasks in this level require readers to perform single, relatively simple arithmetic operations, such as addition. The numbers to be used are provided and the arithmetic operation to be performed is specified.
Level 2 *226-275*	Some tasks in this level require readers to locate a single piece of information in the text; however, several distractors or plausible but incorrect pieces of information may be present or low-level inferences may be required. Other tasks require the reader to integrate two or more pieces of information or to compare and contrast easily identifiable information based on a criterion provided in the question or directive.	Tasks in this level are more varied than those in level 1. Some require the readers to match a single piece of information; however, several distractors may be present or the match may require low-level inferences. Tasks in this level may also ask the reader to cycle through information in a document or to integrate information from various parts of a document.	Tasks in this level typically require readers to perform a single operation using numbers that are either stated in the task or easily located in the material. The operation to be performed may be stated in the question or easily determined from the format of the material (for example, an order form).
Level 3 *276-325*	Tasks in this level tend to require readers to make literal or synonymous matches between the text and information given in the tasks, or to make matches that require low-level inferences. Other tasks ask readers to integrate information from dense or lengthy text that contains no organizational aids such as headings. Readers may also be asked to generate a response based on information that can be easily identified in the text. Distracting information is present, but is not located near the correct information.	Some tasks in this level require the reader to integrate multiple pieces of information from one or more documents. Others ask readers to cycle through rather complex tables or graphs which contain information that is irrelevant or inappropriate to the task.	In tasks in this level, two or more numbers are typically needed to solve the problem, and these must be found in the material. The operation(s) needed can be determined from the arithmetic relation terms used in the question or directive.
Level 4 *326-375*	These tasks require readers to perform multiple-feature matches and to integrate or synthesize information from complex or lengthy passages. More complex inferences are needed to perform successfully. Conditional information is frequently present in tasks at this level and must be taken into consideration by the reader.	Tasks in this level, like those at the previous levels, ask readers to perform multiple-feature matches, cycle through documents, and integrate information; however, they require a greater degree of inferencing. Many of these tasks require readers to provide numerous responses but do not designate how many responses are needed. Conditional information is also present in the document tasks at this level and must be taken into account by the reader.	These tasks tend to require readers to perform two or more sequential operations or a single operation in which the quantities are found in different types of displays, or the operations must be inferred from semantic information given or drawn from prior knowledge.
Level 5 *376-500*	Some tasks in this level require the reader to search for information in dense text which contains a number of plausible distractors. Others ask readers to make high-level inferences or use specialized background knowledge. Some tasks ask readers to contrast complex information.	Tasks in this level require the reader to search through complex displays that contain multiple distractors, to make high-level text-based inferences, and to use specialized knowledge.	These tasks require readers to perform multiple operations sequentially. They must disembed the features of the problem from text or rely on background knowledge to determine the quantities or operations needed.

BEST COPY AVAILABLE

7.5 Conclusions

The definition of literacy and the measurement framework used in the IALS were developed and refined in three large-scale adult literacy surveys conducted in the United States. Using data from these earlier assessments, researchers identified variables which underlie successful performance on a broad array of literacy tasks. These variables have been useful in developing new literacy tasks and in extending and refining literacy measurement. Moreover, they have provided a framework for understanding and articulating what is being measured. As a result, it is possible to interpret the survey results in terms of levels of performance that are both generalizable and valid across time and across population groups.

Chapter 8:

Validity Generalization of the Assessment Across Countries

Don Rock

8.1 Introduction

This chapter reports on an empirical investigation of the discriminant validity of the three constructs assumed to underlie the International Adult Literacy Survey assessment. Specifically, the goal of the investigation was to determine the relative independence of the three adult literacy scales (prose, document, and quantitative) and the way in which this independence may vary from country to country. The three scales are expected to be highly related since they all require general reading skills. There should also be some unique variance associated with each scale, however, because individuals' reading performance is at least partly a function of their prior knowledge and experience with the type of content matter reflected in the literacy tasks. Adults with academic training may perform especially well on the prose scale, for example, while individuals whose day-to-day work requires them to read and interpret technical documents may do best on the document scale.

The analyses undertaken were not concerned with mean differences in literacy levels between countries. Rather, the purpose of the analyses was to answer two questions:

- Can the observed relationships between the three constructs assumed to underlie the three literacy scales be efficiently described with just a single underlying general literacy factor, or do the constructs indeed represent three different literacy factors, albeit with a large shared general literacy component?
- Does the pattern and/or level of the correlations among the factors remain invariant across countries, or do the interrelationships

among the three types of literacy vary systematically country by country? For example, are the relationships between the prose and document factors lower in countries with educational systems that tend to "track" students into academic versus vocational pursuits relatively early in their educational sequence, relative to countries that do little or no specialized tracking?

8.2 Sample and Methods

Data from all of the IALS countries were included in this analysis. The relevant sample sizes are presented in Table 8.1.

Table 8.1: **Study sample size, by country/language group**

Country	Sample size for analysis
Canada/English	3,343
Canada/French	1,333
France	2,530
Germany	1,980
Netherlands	3,017
Poland	2,238
Sweden	2,920
Switzerland/French	1,254
Switzerland/German	1,194
United States	2,138

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

As described in Chapter 2 of this report, the data from each country were based on probability samples.

Confirmatory factor analysis (LISREL8; Jöreskog and Sörbom 1993) was used to test two competing models. The first, a single-factor model, hypothesized that a single general literacy factor could explain the correlations between the three literacy scales. Within each country, two separate scores were generated for each of the three scales by scoring the odd and even items separately. Thus, each scale was represented by two relatively comparable scores. The six scores generated by the odd-even halves served as the input for all succeeding analyses. The single-factor model assumed that a single factor could satisfactorily reproduce the six-by-six covariance matrices in each country.

The second model was a three-factor model, with each pair of odd-even scores defining a separate factor. The question here is: does the three-factor model provide a significant improvement over the single-factor model and, if so, how do the correlations among the factors vary by country?

It should be noted that, unlike the results presented elsewhere in this report, the analyses reported here do not depend on the scaling model or its assumptions. The scaling model is, of course, somewhat more model-dependent, in that it makes certain assumptions about the shape of the functional relationship between item performance and the literacy scores (thetas). The extent to which the factor correlations based on simple number correct scoring lead to similar results with respect to the inter-scale correlations provides additional evidence for the validity of the more model-dependent scaling assumptions.

8.3 Results

The overall chi-square for the single-factor solution was 3917.08 with 90 degrees of freedom (p=0.00), suggesting a relatively poor fit. The finding of a large and significant chi-square was not surprising, however, because the total sample size is equal to the sum of the individual sample n's, and the overall chi-square which measures lack of fit across all samples combined is a multiplicative function of the sample n's. A more appropriate comparison would be to contrast the single-factor model goodness-of-fit indices with those of the less parsimonious three-factor solution. Table 8.2 presents selected goodness-of-fit indices for the two models.

Table 8.2: **Selected goodness of fit indices for single-factor and three-factor models[a]**

Models	x^2/degrees of freedom	Root mean square standard error analysis[b]	Standardized root mean square residual	Non-normed fix index[c]	Critical sample size[d]
Single factor	43.74	.04	.04	.94	692.38
Three factor	3.18	.01	.01	1.00	10,157.73

[a] This table contains results derived from all countries pooled into a common dataset.

[b] RMSEA (root mean square standard error analysis) is a measure of discrepancy per degree of freedom that takes into consideration that the model only approximately fits in the population. RMSEA of less than .05 are considered good fits.

[c] The non-normed fit index is the Tucker-Lewis index, which varies between 0 and 1 and is often interpreted as a measure of the reliability of the model.

[d] The critical sample size indicates the size of the sample needed to reject the fit of the hypothesized model.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Inspection of the goodness of fit indices in Table 8.2 suggests that there is a considerable reduction in the chi-square to degrees of freedom ratios (by a factor of about 15:1) as one moves from the simpler single-factor model to the more complex three-factor model. Clearly, the three-factor solution seems to be a significant improvement over the single general literacy factor solution.

Tables 8.3 to 8.22 present the standardized factor loadings, factor intercorrelations, and selected goodness-of-fit indices for the ten populations in the analysis.

Table 8.3: **Factor loadings: Canada/English**

	Prose	Document	Quantitative
Prose Odd	.86 (.01)	0	0
Prose Even	.91 (.01)	0	0
Document Odd	0	.89 (.01)	0
Document Even	0	.90 (.01)	0
Quantitative Odd	0	0	.88 (.01)
Quantitative Even	0	0	.89 (.01)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

Table 8.4: **Factor intercorrelations: Canada/English**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.92 (.01)	1.00	-
Quantitative	.89 (.01)	.94 (.01)	1.00

Percentage contribution to overall chi-square = 16.8

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.5: **Factor loadings: Canada/French**

	Prose	Document	Quantitative
Prose Odd	.87 (.02)	0	0
Prose Even	.89 (.02)	0	0
Document Odd	0	.91 (.02)	0
Document Even	0	.92 (.02)	0
Quantitative Odd	0	0	.87 (.02)
Quantitative Even	0	0	.90 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8. 6: **Factor intercorrelations: Canada/French**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.93 (.01)	1.00	-
Quantitative	.91 (.01)	.95 (.01)	1.00

Percentage contribution to overall chi-square = 31.1

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.7: **Factor loadings: France**

	Prose	Document	Quantitative
Prose Odd	.86 (.02)	0	0
Prose Even	.88 (.02)	0	0
Document Odd	0	.90 (.02)	0
Document Even	0	.91 (.02)	0
Quantitative Odd	0	0	.90 (.02)
Quantitative Even	0	0	.90 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.8: **Factor intercorrelations: France**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.92 (.01)	1.00	-
Quantitative	.89 (.01)	.94 (.01)	1.00

Percentage contribution to overall chi-square = 4.9

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.9: **Factor loadings: Germany**

	Prose	Document	Quantitative
Prose Odd	.86 (.02)	0	0
Prose Even	.84 (.02)	0	0
Document Odd	0	.88 (.02)	0
Document Even	0	.89 (.02)	0
Quantitative Odd	0	0	.84 (.02)
Quantitative Even	0	0	.87 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.10: **Factor intercorrelations: Germany**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.90 (.01)	1.00	-
Quantitative	.88 (.01)	.92 (.01)	1.00

Percentage contribution to overall chi-square = 6.4

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.11: **Factor loadings: Netherlands**

	Prose	Document	Quantitative
Prose Odd	.84 (.02)	0	0
Prose Even	.87 (.02)	0	0
Document Odd	0	.86 (.02)	0
Document Even	0	.86 (.01)	0
Quantitative Odd	0	0	.85 (.02)
Quantitative Even	0	0	.87 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.12: **Factor intercorrelations: Netherlands**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.89 (.01)	1.00	-
Quantitative	.84 (.01)	.94 (.01)	1.00

Percentage contribution to overall chi-square = 1.5

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.13: **Factor loadings: Poland**

	Prose	Document	Quantitative
Prose Odd	.88 (.02)	0	0
Prose Even	.88 (.02)	0	0
Document Odd	0	.90 (.02)	0
Document Even	0	.91 (.02)	0
Quantitative Odd	0	0	.89 (.02)
Quantitative Even	0	0	.89 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.14: **Factor intercorrelations: Poland**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.90 (.01)	1.00	-
Quantitative	.86 (.01)	.94 (.01)	1.00

Percentage contribution to overall chi-square = 6.7

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.15: **Factor loadings: Sweden**

	Prose	Document	Quantitative
Prose Odd	.87 (.02)	0	0
Prose Even	.88 (.02)	0	0
Document Odd	0	.87 (.01)	0
Document Even	0	.89 (.01)	0
Quantitative Odd	0	0	.88 (.02)
Quantitative Even	0	0	.87 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.16: **Factor intercorrelations: Sweden**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.88 (.01)	1.00	-
Quantitative	.83 (.01)	.94 (.01)	1.00

Percentage contribution to overall chi-square = 5.8

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.17: **Factor loadings: Switzerland/French**

	Prose	Document	Quantitative
Prose Odd	.80 (.02)	0	0
Prose Even	.87 (.02)	0	0
Document Odd	0	.87 (.02)	0
Document Even	0	.88 (.02)	0
Quantitative Odd	0	0	.83 (.02)
Quantitative Even	0	0	.87 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.18: **Factor intercorrelations: Switzerland/French**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.87 (.01)	1.00	-
Quantitative	.83 (.02)	.91 (.01)	1.00

Percentage contribution to overall chi-square = 7.0

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.19: **Factor loadings: Switzerland/German**

	Prose	Document	Quantitative
Prose Odd	.84 (.03)	0	0
Prose Even	.85 (.02)	0	0
Document Odd	0	.91 (.02)	0
Document Even	0	.90 (.02)	0
Quantitative Odd	0	0	.84 (.02)
Quantitative Even	0	0	.85 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.20: **Factor intercorrelations: Switzerland/German**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.83 (.01)	1.00	-
Quantitative	.82 (.02)	.92 (.01)	1.00

Percentage contribution to overall chi-square = 13.5

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.21: **Factor loadings: United States**

	Prose	Document	Quantitative
Prose Odd	.85 (.02)	0	0
Prose Even	.89 (.02)	0	0
Document Odd	0	.90 (.02)	0
Document Even	0	.90 (.02)	0
Quantitative Odd	0	0	.87 (.02)
Quantitative Even	0	0	.84 (.02)

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 8.22: **Factor intercorrelations: United States**

	Prose	Document	Quantitative
Prose	1.00	-	-
Document	.87 (.01)	1.00	-
Quantitative	.88 (.01)	.91 (.01)	1.00

Percentage contribution to overall chi-square = 6.3

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

This confirmatory solution only constrained the number of factors and the pattern of loadings to be the same across all countries. This allowed the correlations between factors and the size of the actual loadings to differ in each population (country). Inspection of the standardized loadings (i.e., the correlations between the observed scale odd-even scores and the underlying factors) suggests that the pattern of salient loading did not differ much across countries with certain minor exceptions. In most countries, the document indicators tended to have as high or slightly higher loadings than did the indicators of the other two scales. This suggests that the reliability of the document scale may be somewhat higher than that of the other scales, because it is a slightly longer scale. The factor correlations among the three scales are in the high 80s or low 90s across most of the countries. In general, the lowest correlations (.82 to .91) are between the prose and quantitative scales, while the highest correlations are between the document and quantitative scales (.91 to .95).

The factor intercorrelations are relatively high on average. The intercorrelations between the document and quantitative scales are especially high, but even these are significantly less than 1.0, indicating that some distinct literacy abilities are captured by each scale. It should be kept in mind that the factor intercorrelations from this confirmatory solution are true score correlations and thus can be interpreted as estimates of the unattenuated correlations among the factors. It is interesting to note that, on average, the factor intercorrelations in the IALS are consistently lower than those found in the U.S. National Adult Literacy Survey (NALS) sample. The NALS sample may have been more heterogeneous than the present sample, however; this would inflate the factor intercorrelations.

8.4 Conclusions

The analyses described here, based on 10 IALS populations, tend to replicate the findings of earlier studies, which have indicated that the three literacy scales (prose, document, and quantitative) represent three separate but relatively highly correlated aspects of literacy. Although a strong general literacy factor was found in all 10 populations, there was sufficient separation among the three literacy scales to justify reporting these scales separately. The document and quantitative scales tended to be the most highly correlated, but in no case did they converge into a single factor.

Chapter 9:

An Analysis of Items with Different Parameters Across Countries

Marilyn R. Binkley and Jean R. Pignal

9.1 Introduction

In international comparative studies, one central concern is whether the assessment items measure the same latent trait across countries, language groups, and cultural communities. Historically, one way to test this concern has been to have item texts translated from the common language into a target language and then back into the common language by two independent translators. If there is a reasonably close match between the original and translated versions, one can assume that the items are identical, at least on the surface, and that they measure the same latent trait.

Another way to evaluate whether items are measuring the same latent trait is to evaluate whether the items perform in the same way across countries. Comparable performance implies that people with the same level of ability would be likely to get the same items right or wrong.

Comparative studies designed to measure the achievement of the school population in academic subjects have also sought to determine whether students across countries have had equal opportunity to learn the appropriate subject matter. Teacher questionnaires which focus on students' opportunity to learn have proven especially useful in this area of inquiry.

The IALS study presented a different challenge. Here, the goal was to estimate the literacy proficiencies of adults in nations in which languages and cultures differed. Literacy is known to have multiple antecedents in addition to schooling, and these might be directly traced to cultural variations that influence the acquisition of literacy skills. Because the survey targeted adults aged 16 to 65, ties to school curricula were limited. Instead, adults' life experiences would play a major role in the results, because these experiences influence the acquisition and maintenance of literacy skills.

Although background variables could partially account for cultural variations in literacy proficiencies, it was important to establish whether the latent trait—that is, the underlying processes that characterize literacy—was universal and transcended linguistic and cultural boundaries. The IALS provided a valuable opportunity to study this issue using the theory of literacy processes that served as the framework for the assessment (see Chapter 7).

The following section briefly summarizes the framework used to develop the assessment items and to explain item difficulty. Subsequent sections analyze differences between the French and English versions of particular items and relate these differences to variations in item characteristics across countries.

9.2 A Theory of the Cognitive Processes Underlying Literacy Proficiency

Based on their work on the Young Adult Literacy Study (1986), the U.S. Department of Labor literacy study (1992), and the National Adult Literacy Survey (1993), Kirsch and Mosenthal found that most adults have some degree of literacy proficiency, and that those with greater proficiency have a higher probability of performing more difficult tasks correctly than those with lower proficiency. To understand literacy proficiency, they argued, one needed to understand the nature of tasks which individuals with known literacy proficiencies are likely to perform successfully. Their work showed that tasks with similar processing characteristics tend to cluster within definable ranges of difficulty (Kirsch, Jungeblut, Jenkins, and Kolstad 1993). Kirsch and Mosenthal then systematically analyzed the interaction between items and literacy proficiency, isolating and describing the cognitive processes which underlie literacy task difficulty. They identified two types of variables that they believed are central to literacy task difficulty: *task characteristics* and *material characteristics*.

The category of *task characteristics* included the following elements.

- *Type of match*: How explicit is the relationship between what the respondent is asked to do and the presentation of the required information in the text? Do individuals merely need to locate the information? Or do they need to cycle through the text to identify two or more pieces of conditional information, compare or contrast pieces of information, or use special knowledge?
- *Plausibility of distractors*: Is there information in the text that appears to be, but is not, the correct answer?
- *Type of information*: Is the requested information concrete (e.g., the name of a person or thing) or abstract (e.g., reasons for an interpretation)?
- *Operation specificity*: How easy or difficult is it to set up the quantitative problem?
- *Type of calculation*: Does a quantitative problem involve addition, subtraction, multiplication, division, or some combination of these operations?

- *Structural complexity*: Is the information organized as a simple list or a combined list? Does it contain an underlying intersected list or nested list? Is it composed of different multiple documents?

The materials on which the assessment tasks were based included tables, indexes, lists, coupons, schedules, charts, graphs, maps, and forms. *Characteristics of materials* included:

- *Number of syllables per hundred words of text*: Are the words in the text short or long?
- *Number of sentences per hundred words of text*: Does the text contain complex sentences with many embedded clauses?
- *Number of labels*. Are documents clearly structured and labeled?
- *Type of document*: Is a document a simple list, or a combined, intersected or nested list? Does it contain more than one structure?
- *Amount of information*. How much information must the reader search for in the text?

Based on the research literature and empirical evidence, Kirsch and Mosenthal developed a ranking system using these variables (see Chapter 7). Item characteristics reflecting difficulty were used as dependent measures in regression models to see how much of the variance in literacy proficiencies was accounted for by these variables.

Kirsch and Mosenthal found that for both prose and document literacy, the variables most highly related to task difficulty were type of match, plausibility of distractors, and type of information. Readability, as measured by number of syllables and number of sentences per hundred words of text, was less significant. In the area of quantitative literacy, they found that operation specificity, type of calculation, and plausibility of distractors were most strongly related to task difficulty. Type of match and readability were of little consequence. Overall, the empirical evidence suggested that differences in the complexity of cognitive processing inherent in literacy tasks were directly related to specific skills.

These findings have interesting implications for international comparative studies. It is striking to see, for example, that readability had comparatively little impact on the difficulty of the literacy assessment items. From the standpoint of international assessments, this might indicate that syntactical differences across languages would be unlikely to increase or decrease the difficulty of literacy tasks. One could therefore hypothesize that maintaining the semantic structure of a piece of text through translation would result in assessment items of equal difficulty.

Further, the fact that Kirsch and Mosenthal found analogous skills in the areas of prose and document literacy (and, to a lesser extent, quantitative literacy) suggests the existence of a less context-specific, more generalizable set of literacy skills, transcending language and cultural boundaries. In other words, the differences in symbolic systems may be less important than expected. On the other hand, the universality of the literacy tasks that adults perform, and the concreteness of the stimulus materials and task directives, may be particularly important.

9.3 Application of the Theory to the IALS

As described in Chapter 7, items developed for the IALS were based on the framework used in three previous large-scale assessments: the Young Adult Literacy Survey (YALS), the U.S. Department of Labor (DOL) survey, and the National Adult Literacy Survey (NALS). As a result, the IALS items shared the same characteristics as the items in these earlier surveys. The English versions of the IALS items were reviewed and tested to determine whether they fit into the literacy scales in accordance with the theory and whether they were consistent with the NALS data. Quality control procedures for item translation, scoring, and scaling followed the same procedures used in the NALS and extended the methods used in other international studies (see Chapters 5 and 10).

Most of the items administered in the IALS were successful from a psychometric standpoint. Consistent with national and international experiences, however, and despite stringent efforts at quality control, some of the assessment items did not meet the criteria for inclusion in the final tabulation of results (see Chapter 10). Specifically, in carrying out the IRT modeling used to create the three literacy scales, researchers found that a number of assessment items had significantly different item parameters across the IALS countries.

In response to this statistical evidence, 13 items were dropped from the scales. Twelve of the items were eliminated because the item characteristics differed significantly in more than three countries, perhaps due to design flaws or typographical errors. Two items on the quantitative scale proved to be perfectly correlated, hence the second item offered no information value to the IRT estimation models. The second item was, therefore, dropped. Another 46 items were given unique item parameters for three or fewer countries. These items offer a naturally occurring experiment within the confines of the larger survey. Studying these items would, it was hoped, illuminate differences in cognitive processes within these relatively constrained situations. The goal was to see if it was possible to predict whether an item would be more or less difficult in particular countries given differences in the translations, and to see if these differences were related to the framework underlying the IALS item development.

The researchers conducting this study were proficient in English and French only, so the available pool of 46 items was reduced to 14 items that had been assigned unique parameters in both English-speaking countries (the United States and Canada) or in one of the French-speaking countries (Canada, France, or Switzerland). Because the United States and Canada used identical item formats, there was no possibility of explaining differences between those countries based on the stimulus text or directive alone. Factors other than the instrument would have to be considered. On the other hand, the three French-language tests were unique translations, so differences among the three could be attributed to differences in "dialect" or country-specific community attributes.

The 14 items with unique item parameters were carefully analyzed to determine whether there was a reasonable explanation for differences in item performance across the countries. Seven of these items are discussed in the following section. These seven examples illustrate the impact of what might appear to be minor, idiosyncratic variations among the different versions of each item.

9.4 Analysis of Items

As explained earlier in this chapter, researchers inspected the different versions of each item chosen for analysis to see whether there were differences across the translations in the type of match or type of information requested. If differences were found, then the goal was to predict the direction of the impact based on Kirsch and Mosenthal's hierarchy for ranking item difficulty. This prediction could then be tested against the item characteristic curves.

Example 1

This example is based on a weather map that supposedly appeared in a newspaper. The respondent is asked to circle the area where heavy rain is expected through the weekend.

Canada

Question 8. Sur la carte météorologique de l'Europe, encerclez la région où l'on prévoit de fortes pluies durant la fin de semaine.

BULLETIN MÉTÉOROLOGIQUE

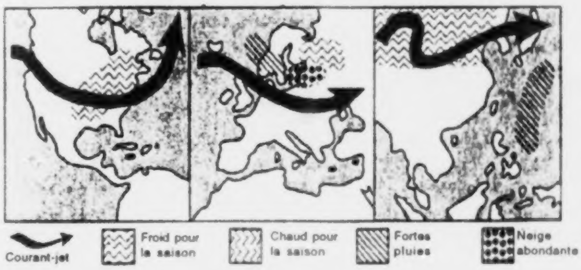
Europe

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Algarve	19	7	e	21	48	e
Amsterdam	11	6	np	12	44	np
Ankara	17	7	np	19	46	np
Athènes	22	15	np	23	57	np
Barcelone	16	8	e	14	48	e
Belgrade	14	6	np	10	34	n
Berlin	8	2	n	6	34	n
Bruxelles	11	6	np	14	44	np
Budapest	9	1	np	9	35	n
Copenhague	7	1	a	6	35	n
Costa del Sol	21	8	e	21	50	e
Dublin	10	6	np	13	46	np
Édimbourg	10	6	n	10	43	n
Florence	11	5	e	14	43	e
Francfort	12	6	np	13	39	np
Genève	9	2	e	12	39	e
Helsinki	-1	-7	m	-3	15	np
Istanbul	17	10	np	15	48	a
Las Palmas	26	18	np	27	64	np
Lisbonne	19	9	e	19	50	e
Londres	12	5	np	13	44	np
Madrid	17	3	e	18	39	e
Milan	9	3	e	13	43	e
Moscou	1	-3	a	-3	13	m
Munich	11	3	np	12	43	np
Nice	14	7	e	15	46	e
Oslo	4	-4	n	5	29	n
Paris	12	6	np	13	43	np
Prague	11	1	np	8	35	n
Reykjavik	4	2	p	6	31	n
Rome	20	12	e	20	50	e
Saint-Pétersbourg	-1	-7	m	-4	11	np
Stockholm	1	-5	n	-2	20	n
Strasbourg	12	5	np	15	44	np
Tallinn	-1	-7	m	-4	15	np
Venise	10	3	e	11	39	e
Vienne	9	-1	np	10	35	n
Varsovie	8	2	a	6	34	n
Zurich	8	0	e	9	34	np

Océanie

Auckland	20	14	e	17	11	a
Sydney	27	17	np	25	16	np

Prévisions pour vendredi jusqu'à dimanche



Amérique du Nord
Un courant froid envahira le centre-ouest et le nord-est des États-Unis vendredi et toute la fin de semaine. En dépit du froid qui sévira à Chicago, à Toronto et à New York, le temps sera sec. Des percées de soleil baigneront Los Angeles tous les jours et la température y sera normale pour la saison.

Europe
Une vague de temps doux atteindra l'ouest et le centre de l'Europe vendredi et persistera toute la fin de semaine. Londres et Paris connaîtront un temps sec et quelques percées de soleil de vendredi à dimanche. La pluie continuera d'arroser le sud-ouest de la Norvège. Il neigera sur la région située entre Minsk et Moscou.

Asie
Le typhon Elsie continuera probablement de hanter l'est des Philippines et le sud du Japon dans la journée de vendredi et durant toute la fin de semaine. Risque d'averses et même d'un peu de verglas ou de neige à Séoul. Un courant froid envahira Beijing, d'où possibilités de neige. À Hong Kong, le début de la fin de semaine sera chaud.

Moyen-Orient

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Beyrouth	28	19	np	29	20	e
Le Caire	29	20	np	28	19	np
Damas	24	12	e	26	14	e
Jérusalem	27	15	e	26	14	e
Riyad	34	13	e	32	13	e

Amerique latine

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Buenos Aires	23	11	np	26	13	e
Caracas	29	20	e	31	18	e
Lima	23	17	n	23	16	n
Mexico	23	11	a	23	12	np
Rio de Janiero	32	22	e	28	21	a
Santiago	24	4	e	22	6	np

Légende e-ensoleillé, np-nébulosité partielle, n-nuageux, a-averses, o-orages, p-pluie, m-rafales de neige, n-neige, v-verglas, T-temp. Cartes, prévisions et données fournies par Accu-Weather, Inc. © 1992

Asie

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Bangkok	32	22	np	30	23	e
Beijing	11	0	e	8	2	np
Hong Kong	30	23	e	29	22	np
Manille	31	25	e	31	25	a
New Delhi	31	13	e	32	16	e
Séoul	14	6	np	14	4	np
Shanghai	22	10	np	24	12	e
Singapour	31	24	np	28	23	a
T'ai-pei	26	21	np	26	19	np
Tokyo	18	9	np	17	7	np

Afrique

Alger	27	14	e	26	13	e
Le Cap	20	11	a	18	11	np
Casablanca	20	14	n	21	11	np
Harare	34	17	e	32	18	np
Lagos	30	24	np	29	24	np
Nairobi	27	12	np	26	13	np
Tunis	27	17	np	17	14	np

Amérique du Nord

Anchorage	0	-2	n	3	0	a
Atlanta	14	4	np	8	2	np
Boston	15	4	n	9	-1	np
Chicago	2	-5	n	-2	-8	np
Denver	8	-3	np	4	-6	n
Detroit	4	-2	n	4	-5	np
Honolulu	31	20	e	31	21	np
Houston	15	3	np	12	6	np
Los Angeles	28	14	e	24	13	e
Miami	30	22	np	29	21	np
Minneapolis	-1	-8	n	1	-7	np
Montréal	7	-2	m	4	-3	n
Nassau	31	22	np	28	21	a
New York	14	4	p	10	2	np
Phoenix	23	11	np	22	8	e
San Fran.	20	11	np	21	8	e
Seattle	11	6	np	15	7	p
Toronto	6	-3	n	3	-3	n
Washington	14	6	p	11	4	np

BEST COPY AVAILABLE

France

Question 8. Sur la carte météo de l'Europe, entourez l'endroit où de fortes précipitations sont prévues dans la semaine.

LE TEMPS

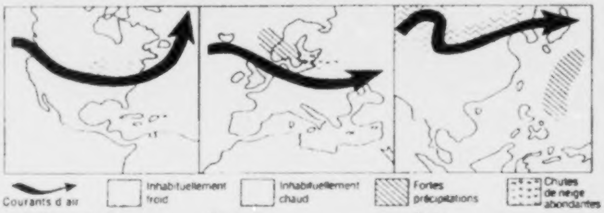
Europe

	Aujourd'hui			Demain		
	Max °C	Min °C	T	Max °C	Min °C	T
Algarve	19	7	s	21	8	s
Amsterdam	11	6	pc	12	7	pc
Ankara	17	7	pc	19	8	pc
Athènes	22	15	pc	23	14	pc
Barcelone	16	8	s	14	9	s
Belgrade	14	6	pc	10	1	c
Berlin	8	2	c	6	1	c
Bruxelles	11	6	pc	14	7	pc
Budapest	9	1	pc	9	2	c
Copenhague	7	1	p	6	2	c
Costa del Sol	21	8	s	21	10	s
Dublin	10	6	pc	13	8	pc
Edimbourg	10	6	c	10	6	c
Florence	11	5	s	14	6	s
Francfort	12	6	pc	13	4	pc
Genève	9	2	s	12	4	s
Helsinki	-1	-7	tn	-3	-10	pc
Istambul	17	10	pc	15	9	a
Las Palmas	26	18	pc	27	18	pc
Lisbonne	19	9	s	19	10	s
Londres	12	5	pc	13	7	pc
Madrid	17	3	s	18	4	s
Milan	9	3	s	13	6	s
Moscou	1	-3	p	-3	-11	tn
Munich	11	3	pc	12	6	pc
Nice	14	7	s	15	8	s
Oslo	4	-4	c	5	-2	c
Paris	12	6	pc	13	6	pc
Prague	11	1	pc	8	2	c
Reykjavik	4	2	p	6	-1	c
Rome	20	12	s	20	10	s
Saint Petersbourg	-1	-7	tn	-4	-12	pc
Stockholm	1	-6	n	-2	-7	c
Strasbourg	12	5	pc	15	7	pc
Tallin	-1	-7	tn	-4	-10	pc
Venise	10	3	s	11	4	s
Vienne	9	1	pc	10	2	c
Varsovie	8	2	a	6	1	c
Zurich	8	0	s	9	1	pc

Océanie

Auckland	20	14	c	17	11	a
Sydney	27	17	pc	25	16	pc

Prévisions de vendredi à dimanche



Amérique du Nord
On attend un temps froid à partir de vendredi et durant tout le week-end sur le Centre-Ouest et le Nord-Est des Etats-Unis. Malgré le froid à Chicago, Toronto et New York City, le temps devrait être sec. Il y aura un peu de soleil et des températures habituelles pour la saison à Los Angeles tout au long du week-end.

Europe
L'Europe de l'Ouest et l'Europe centrale connaîtront une brève période de temps doux à partir de vendredi et durant le week-end. Temps sec et assez ensoleillé durant tout le week-end à Paris et à Londres. Il continuera de pleuvoir à torrent sur le sud-ouest de la Norvège. Chutes de neige entre Minsk et Moscou.

Asie
Le typhon Elsie restera probablement sur l'est des Philippines et le sud du Japon vendredi et durant le week-end. Tendances pluvieuses sur Séoul ; il se pourrait même qu'il y ait un peu de neige ou de verglas. A Pékin, le temps se rafraîchira et il pourrait neiger. Il fera bon à Hong Kong en début de week-end.

Moyen-Orient

	Aujourd'hui			Demain		
	Max °C	Min °C	T	Max °C	Min °C	T
Beyrouth	28	19	pc	29	20	s
Le Caire	29	20	pc	28	19	pc
Damas	24	12	s	26	14	s
Jérusalem	27	15	s	26	14	s
Riyad	34	13	s	32	13	s

Amérique Latine

	Aujourd'hui			Demain		
	Max °C	Min °C	T	Max °C	Min °C	T
Buenos Aires	23	11	pc	26	13	s
Caracas	29	20	s	31	18	s
Lima	23	17	c	23	16	c
Mexico	23	11	a	23	12	pc
Rio de Janeiro	32	22	s	28	21	a
Santiago	24	4	s	22	6	pc

Légende : s – soleil, pc – partiellement couvert, c – couvert, a – averses, o – orages, p – pluie, tn – tempêtes de neige, n – neige, g – gel, T – temps. Cartes, prévisions et données fournies par Accu-Weather, Inc. © 1992

Asie

	Aujourd'hui			Demain		
	Max °C	Min °C	T	Max °C	Min °C	T
Bangkok	32	22	pc	30	23	s
Hong Kong	30	23	s	29	22	pc
Manille	31	25	s	31	25	a
New Delhi	31	13	s	32	16	s
Pékin	11	0	s	8	2	pc
Séoul	14	6	pc	14	4	pc
Shangai	22	10	pc	24	12	s
Singapour	31	24	pc	28	23	a
Taipei	26	21	pc	26	19	pc
Tokyo	18	9	pc	17	7	pc

Afrique

Alger	27	14	s	26	13	s
Casablanca	20	14	c	21	11	pc
Harare	34	17	s	32	18	pc
Lagos	30	24	o	29	24	pc
Le Cap	20	11	a	18	11	pc
Nairobi	27	12	pc	28	13	pc
Tunis	27	17	pc	26	14	pc

Amérique du Nord

Anchorage	0	-2	c	3	0	a
Atlanta	14	4	pc	8	2	pc
Boston	15	4	c	9	-1	pc
Chicago	2	-5	c	-2	-8	pc
Denver	8	-3	pc	4	-6	n
Detroit	4	-2	c	4	-5	pc
Honolulu	31	20	s	31	21	pc
Houston	15	3	pc	12	6	pc
Los Angeles	28	14	s	24	13	s
Miami	30	22	pc	29	21	pc
Minneapolis	-1	-8	c	1	-7	pc
Montréal	7	-2	tn	4	-3	c
Nassau	31	22	pc	28	21	a
New York	14	4	p	10	2	pc
Phoenix	23	11	pc	22	8	s
San Francisco	20	11	pc	21	8	s
Seattle	11	6	pc	13	7	p
Toronto	6	-3	c	3	-3	c
Washington	14	6	p	11	4	pc

Switzerland

Question B8. Sur la carte météorologique de l'Europe, encerclez la région où l'on prévoit de fortes pluies durant le week-end.

BULLETIN METEOROLOGIQUE

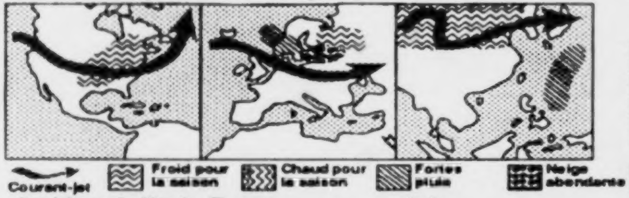
Europe

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Algarve	19	7	e	21	46	e
Amsterdam	11	6	np	12	44	np
Ankara	17	7	np	19	48	np
Athènes	22	15	np	23	57	np
Barcelone	16	8	e	14	48	e
Belgrade	14	6	np	10	34	n
Berlin	8	2	n	6	34	n
Bruxelles	11	6	np	14	44	np
Budapest	9	1	np	9	35	n
Copenhague	7	1	e	6	35	n
Costa del Sol	21	8	e	21	50	e
Dublin	10	6	np	13	48	np
Edimbourg	10	6	n	10	43	n
Florence	11	5	e	14	43	e
Francfort	12	6	np	13	39	np
Genève	9	2	e	12	39	e
Helsinki	-1	-7	rn	-3	15	np
Istanbul	17	10	np	15	48	e
Las Palmas	26	18	np	27	64	np
Lisbonne	19	9	e	19	50	e
Londres	12	5	np	13	44	np
Madrid	17	3	e	18	39	e
Milan	9	3	e	13	43	e
Moscou	1	-3	e	-3	13	rn
Munich	11	3	np	12	43	np
Nice	14	7	e	15	46	e
Oslo	4	-4	n	5	28	n
Paris	12	6	np	13	43	np
Prague	11	1	np	8	35	n
Reykjavik	4	2	p	6	31	n
Rome	20	12	e	20	50	e
Saint Pétersb.	-1	-7	rn	-4	11	np
Stockholm	1	-5	n	-2	20	n
Strasbourg	12	5	np	15	44	np
Tallinn	-1	-7	rn	-4	15	np
Venise	10	3	e	11	38	e
Vienne	9	-1	np	10	35	n
Varsovie	8	2	a	6	34	n
Zurich	8	0	e	9	34	np

Océanie

Auckland	20	14	e	17	11	e
Sydney	27	17	np	25	16	np

Prévisions pour vendredi jusqu'à dimanche



Amérique du Nord
Un courant froid envahira le centre-ouest et le nord-est des Etats-Unis vendredi et pour tout le week-end. En dépit du froid qui sévira à Chicago, à Toronto et à New York, le temps sera sec. Des percées de soleil baigneront Los Angeles tous les jours et la température y sera normale pour la saison.

Europe
Une vague de temps doux atteindra l'ouest et le centre de l'Europe vendredi et persistera tout le week-end. Londres et Paris connaîtront un temps sec et quelques percées de soleil de vendredi à dimanche. La pluie continuera d'arroser le sud-ouest de la Norvège. Il neigera sur la région située entre Minsk et Moscou.

Asie
Le typhon Elsie continuera probablement de hanter l'est des Philippines et le sud du Japon dans la journée de vendredi et durant tout le week-end. Risque d'averses et même d'un peu de verglas ou de neige à Séoul. Un courant froid envahira Pékin, d'où possibilités de neige. A Hong Kong, le début de la fin de semaine sera chaud.

Moyen-Orient

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Beyrouth	28	19	np	29	20	e
Le Caire	29	20	np	28	19	np
Damas	24	12	e	26	14	e
Jérusalem	27	15	e	26	14	e
Riyad	34	13	e	32	13	e

Amérique latine

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Buenos Aires	23	11	np	26	13	e
Caracas	29	20	e	31	18	e
Lima	23	17	n	23	16	n
Mexico	23	11	a	23	12	np
Rio de Janeir	32	22	e	28	21	a
Santiago	24	4	e	22	6	np

Légende: e-ensoleillé, np-nébulosité partielle, n-nuageux, a-averses, o-orages, p-pluie, rn-rafales de neige, n-neige, v-verglas, T-temps. Cartes, prévisions et données fournies par Accu-Weather, Inc. © 1993

Asie

	Aujourd'hui			Demain		
	Max. °C	Min. °C	T	Max. °C	Min. °C	T
Bangkok	32	22	np	30	23	e
Hong Kong	30	23	e	29	22	np
Manille	31	25	e	31	25	a
New Delhi	31	13	e	32	16	e
Pékin	11	0	e	8	2	np
Séoul	14	6	np	14	4	np
Shanghai	22	10	np	24	12	e
Singapour	31	24	np	28	23	a
Taipei	26	21	np	26	19	np
Tokyo	18	9	np	17	7	np

Afrique

Alger	27	14	e	26	13	e
Le Cap	20	11	e	18	11	np
Casablanca	20	14	n	21	11	np
Harare	34	17	e	32	18	np
Lagos	30	24	np	29	24	np
Nairobi	27	12	np	26	13	np
Tunis	27	17	np	17	14	np

Amérique du Nord

Anchorage	0	-2	n	3	0	a
Atlanta	14	4	np	8	2	np
Boston	15	4	n	9	-1	np
Chicago	2	-5	n	-2	-8	np
Denver	8	-3	np	4	-6	n
Detroit	4	-2	n	4	-5	np
Honolulu	31	20	e	31	21	np
Houston	15	3	np	12	6	np
Los Angeles	28	14	e	24	13	e
Miami	30	22	np	29	21	np
Minneapolis	-1	-8	n	1	-7	np
Montréal	7	-2	rn	4	-3	n
Nassau	31	22	np	28	21	a
New York	14	4	p	10	2	np
Phoenix	23	11	np	22	8	e
San Francisco	20	11	np	21	8	e
Seattle	11	6	np	15	7	p
Toronto	6	-3	n	3	-3	n
Washington	14	6	p	11	4	np

BEST COPY AVAILABLE

There are subtle differences among the three versions of this item that appear to be significant. For example, while the words used in the Canadian and Swiss French versions are the same, France's version seems somewhat colloquial or conversational. Further, in France's version the respondent is asked to indicate what will happen during the week, not the weekend. The latter difference may have made France's version of this item somewhat more difficult than the Canadian or Swiss French versions.

Another difference among the versions of this item is found in the map. In Canada's and France's versions, the ocean is left blank. In the Swiss item, however, a dotted pattern is used for the ocean, and these dots create a visual distraction from the diagonal lines which indicate heavy rain. Because the item requires respondents to find and interpret the pattern for heavy rain, the different pattern used for the ocean in the Swiss version is likely to affect the difficulty of the item. We would expect Swiss respondents to find this item more difficult than their French or Canadian counterparts.

Example 2

In this prose item, respondents are asked to read an excerpt from an employment pamphlet titled *The Hiring Interview*. The text is clearly divided into four sections labeled: *Preinterview*, *The Interview*, *The Most Common Types of Interview*, and *Postinterview*. The prompt draws respondents' attention to the final section by asking for the main purpose of the postinterview review.

Canada

Question 9. D'après le dépliant, quel est le but principal de **la revue** après l'entrevue?

Après l'entrevue

Notez les principaux points qui ont été discutés. Comparez les questions qui vous ont causé des difficultés et celles qui vous ont permis de faire valoir vos points forts. **Une telle revue** vous aidera à vous préparer à de futures entrevues. Si vous le désirez, vous pouvez en discuter avec l'agent de placement ou l'orienteur professionnel de votre bureau d'emploi local.

Switzerland

Question C9. D'après le dépliant, quel est l'objectif principal de **la revue** après l'entretien?

Après l'entretien

Notez les points essentiels abordés. Comparez les questions qui vous ont mis en difficulté et celles qui vous ont permis, au contraire, de faire valoir vos points forts. **Une telle revue** vous aidera à vous préparer à de futures entretiens; au besoin, vous pourrez en parler avec le prospecteur-placier ou le conseiller professionnel de votre agence locale.

BEST COPY AVAILABLE

France

Question 9. D'après la brochure, quel est le principal intérêt de **l'après-entretien**?

> Après l'entretien
>
> Notez les points essentiels abordés. Comparez les questions qui vous ont mis en difficulté et celles qui vous ont permis, au contraire, de faire valoir vos points forts. **Ces éléments vous aideront à aborder un nouvel entretien** ; au besoin, vous pourrez en parler avec le prospecteur-placier ou le conseiller professionnel de votre agence locale.

Although the Swiss and Canadian translations differ, both versions include certain key words. Specifically, the prompt asks for the main purpose of the postinterview review, and the text points to "une telle revue." In contrast, France's version does not use any connecting words or phrases beyond those that direct the respondent to the appropriate paragraph. We would therefore expect France's version of this item to be more difficult than the Swiss or Canadian versions because the respondent must infer the correct response from the information provided.

Examples 3 and 4

The next two items are based on a recipe for scrambled eggs with tomatoes, something that would probably be familiar to most adults.

Canada

Question 2. If you want to make enough scrambled eggs for six people, how many eggs should you use?

Question 3. If you decide to make just enough scrambled eggs for two people, how many tablespoons of oil would you need?

Scrambled Eggs with Tomatoes

Ingredients for 4 people:

1 garlic clove
1 onion
3 tablespoons oil
500 grams of fresh red tomatoes or 500 gram can of tomatoes
salt
1 teaspoon sugar
6 eggs

Fry chopped garlic and onion in frying pan with oil until transparent. Add tomatoes that have been peeled and chopped (if they are fresh) or mashed with a fork (if they are canned). Add salt and sugar to lessen the acidity. When the mixture begins to thicken, add the eggs, already beaten, and stir well with a wooden spoon. Cook until eggs are set.

Canada

Question 2. Si vous voulez faire assez d'oeufs brouillés pour six personnes, combien d'oeufs devrez-vous employer?

Question 3. Si vous décidez de faire des oeufs brouillés pour deux personnes seulement, combien de cuillerées à soupe d'huile vous faudra-t-il?

OEUFS BROUILLÉS AUX TOMATES

Ingrédients pour quatre portions
1 gousse d'ail
1 oignon
3 cuillerées à soupe d'huile
500 grammes de tomates rouges fraîches ou 500 grammes de tomates en boîte
sel
1 cuillerée à thé de sucre
6 oeufs

Faire frire l'ail et l'oignon haché dans une poêle jusqu'à ce qu'ils deviennent transparents. Ajouter les tomates, pelées et hachées (si elles sont fraîches) ou réduites en purée à l'aide d'une fourchette (s'il s'agit de tomates en boîte). Ajouter sel et sucre afin d'atténuer l'acidité. Lorsque le mélange commence à épaissir, ajouter les oeufs battus et mélanger bien le tout avec une cuillère en bois. Cuire jusqu'à ce que les oeufs soient bien pris.

France

Question 2. Vous devez faire des oeufs brouillés en quantité suffisante pour six personnes. De combien d'oeufs aurez-vous besoin en suivant cette recette ?

Question 3. Si vous décidez de faire des oeufs brouillés en quantité suffisante pour deux personnes seulement, de combien de cuillères à soupe d'huile aurez-vous besoin ?

Œufs brouillés à la tomate

Proportions pour 4 personnes :
1 gousse d'ail
1 oignon
3 cuillères à soupe d'huile
500 grammes de tomates fraîches bien mûres ou 500 grammes de tomates en boîte
du sel
1 cuillère à café de sucre
6 œufs

Émincez l'ail et les oignons, faites-les revenir dans l'huile jusqu'à ce qu'ils soient transparents. Ajoutez les tomates épluchées et coupées en morceaux (pour les tomates fraîches) ou écrasées en purée (pour les tomates en boîte). Ajoutez du sel et du sucre pour atténuer l'acidité. Lorsque le mélange épaissit, ajoutez les œufs que vous aurez au préalable battus en omelette, et tournez avec une cuillère en bois. Continuez la cuisson jusqu'à ce que les œufs soient bien cuits.

These two items ask the respondent to alter the recipe so that instead of making the indicated amount, the quantities are either enlarged to serve six people or reduced to serve just two. To arrive at the correct answers, the respondent must pay attention to two things in the text: the number of people the recipe is expected to serve, and the number of eggs or the amount of oil to be used. All of this information is provided off to one side of the ingredient list.

The different countries' versions of these items vary in two respects. First, France's version of the prompt specifies that the response should be based on this recipe ("en suivant cette recette"), while the other versions do not mention the recipe at all. Second, each version has a textual idiosyncrasy. In the English version, the required information is located in a list labeled "Ingredients for 4 people;" in the French version, it is labeled "Proportions pour 4 personnes;" and in the Canadian French version, it is labeled "Ingredients pour quatre portions."

Given that cooking eggs is a common occurrence in each of these countries, respondents may be inclined to use their prior knowledge of cooking eggs rather than rely on the recipe given. Because the reminder in France's prompt is likely to focus respondents' attention on what is stated in the recipe rather than on their own knowledge, we would expect the French to have less difficulty with these items than would their counterparts in the other countries. On the other hand, question 3 is likely to be more difficult for the French Canadians than for those in the other countries, for three reasons. First, the heading for the ingredient list does not contain a numeral, but uses the word "quatre" instead. Second, the heading refers to portions instead of people, so the reader must infer that the portion is sufficient or ample for one person. Finally, the directions never mention when to use the oil.

Examples 5 and 6

In this set of items, respondents must complete a form using information listed outside of the form. To answer the items correctly, they must not only fill in each of the cells according to the various categories, but also add the information in each column.

Canada

Question 11. Au travail, vous devez tenir un registre de vos heures de travail et déclarer le total à la fin de chaque semaine sur la fiche illustrée à la page ci-contre. Vous êtes payé selon l'échelle salariale suivante :

Lundi-vendredi	Jusqu'à 8 heures	100%
Lundi-vendredi	Au-delà de 8 heures	125%
Samedi	Jusqu'à 8 heures	125%
Samedi	Au-delà de 8 heures	150%
Dimanche	Chaque heure	150%

La fiche illustrée à la page ci-contre indique vos heures de travail pour la semaine. En vous servant de l'échelle salariale illustrée ci-dessus, remplissez le Sommaire hebdomadaire pour chacune des quatre colonnes.

Code de la feuille de paye ________________ Nom de l'employé ______________________________

Numéro de la période ______________________________ Période du __________ au __________ 19___

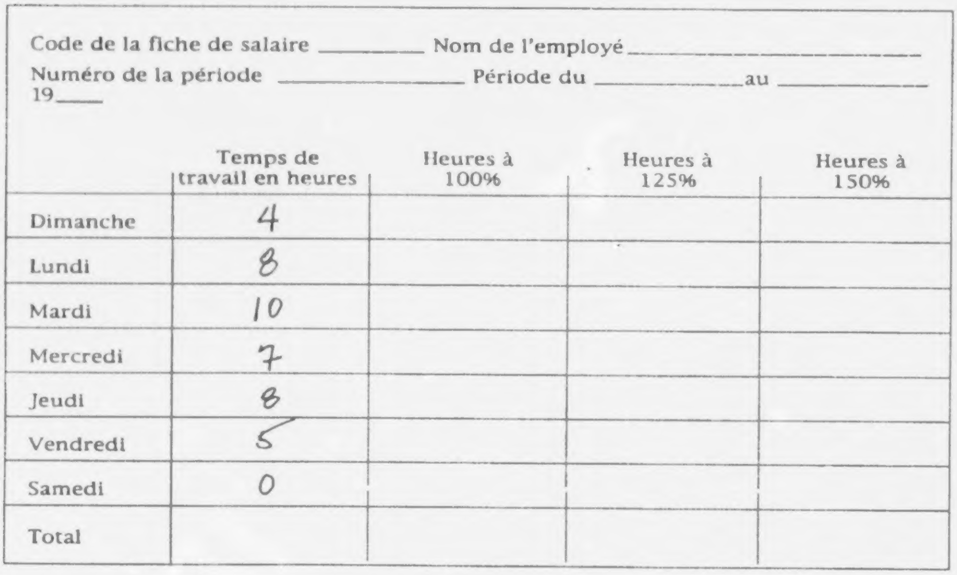
	Heures travaillées	Heures à 100%	Heures à 125%	Heures à 150%
Dimanche	4			
Lundi	8			
Mardi	10			
Mercredi	7			
Jeudi	8			
Vendredi	5			
Samedi	0			
Sommaire hebdomadaire				

Switzerland

Question E11. Au travail, vous devez enregistrer vos heures de travail sur une fiche et déclarer le total à la fin de chaque semaine. Vous êtes payé selon l'échelle de taux suivante :

Lundi-vendredi	Jusqu'à 8 heures	100%
Lundi-vendredi	Au-delà de 8 heures	125%
Samedi	Jusqu'à 8 heures	125%
Samedi	Au-delà de 8 heures	150%
Dimanche	Chaque heure	150%

La fiche illustrée à la page ci-contre indique vos heures de travail pour la semaine. En vous servant de l'échelle salariale illustrée ci-dessus, remplissez le sommaire hebdomadaire pour chacune des quatres colonnes.



Code de la fiche de salaire ________ Nom de l'employé ________________________
Numéro de la période ______________ Période du ____________au ____________
19____

	Temps de travail en heures	Heures à 100%	Heures à 125%	Heures à 150%
Dimanche	4			
Lundi	8			
Mardi	10			
Mercredi	7			
Jeudi	8			
Vendredi	5			
Samedi	0			
Total				

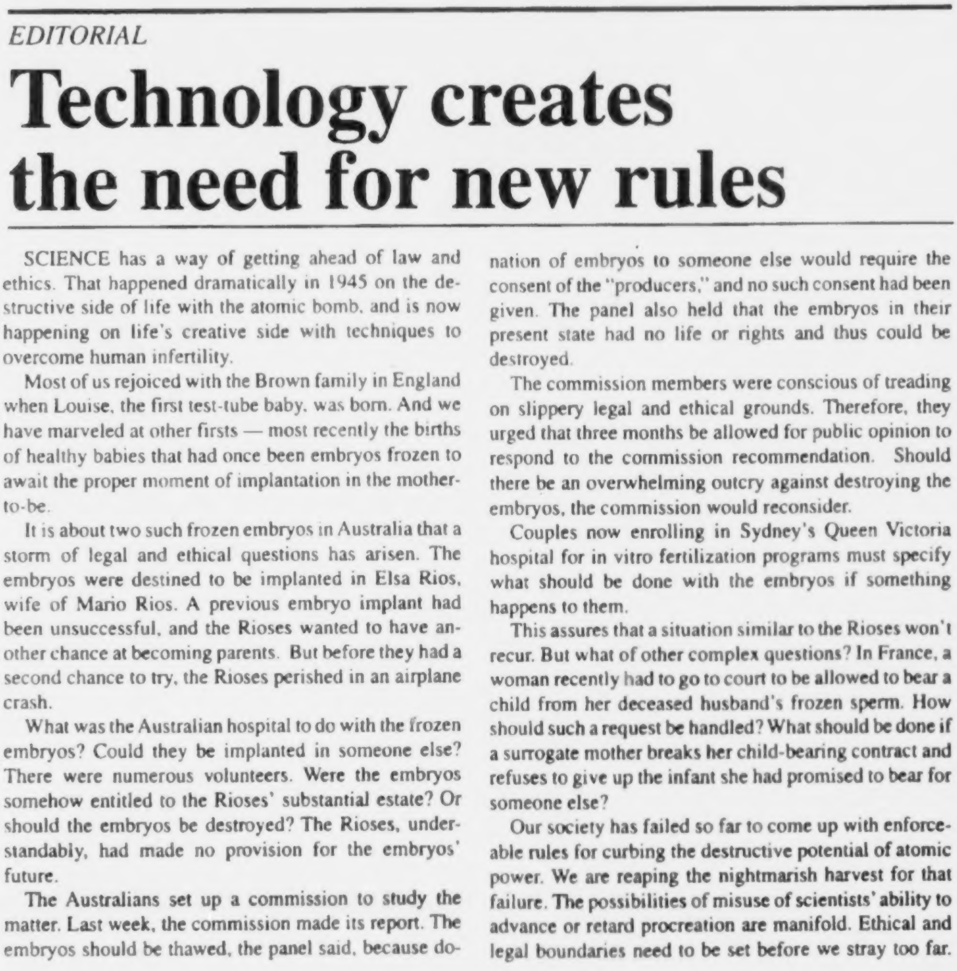
There are two important differences between the forms. The French Canadian version calls for a "sommaire hebdomadaire," and the associated row is boxed twice. The Swiss version, on the other hand, calls for a "total," and the row is not boxed. These two differences are likely to give an advantage to Canadian French respondents (and to United States and Canadian English respondents), because the prompt specifically asks respondents to complete the Weekly Summary. The Swiss version is ambiguous because the form does not include anything marked "sommaire hebdomadaire" to correspond to the prompt.

Example 7

The final example is more complex than the previous examples. This item is based on an editorial discussing technological advances in the preservation of embryos.

Canada

Question 6. Underline the sentence that explains what the Australians did to help decide how to deal with the frozen embryos belonging to a couple killed in the plane crash.



EDITORIAL

Technology creates the need for new rules

SCIENCE has a way of getting ahead of law and ethics. That happened dramatically in 1945 on the destructive side of life with the atomic bomb, and is now happening on life's creative side with techniques to overcome human infertility.

Most of us rejoiced with the Brown family in England when Louise, the first test-tube baby, was born. And we have marveled at other firsts — most recently the births of healthy babies that had once been embryos frozen to await the proper moment of implantation in the mother-to-be.

It is about two such frozen embryos in Australia that a storm of legal and ethical questions has arisen. The embryos were destined to be implanted in Elsa Rios, wife of Mario Rios. A previous embryo implant had been unsuccessful, and the Rioses wanted to have another chance at becoming parents. But before they had a second chance to try, the Rioses perished in an airplane crash.

What was the Australian hospital to do with the frozen embryos? Could they be implanted in someone else? There were numerous volunteers. Were the embryos somehow entitled to the Rioses' substantial estate? Or should the embryos be destroyed? The Rioses, understandably, had made no provision for the embryos' future.

The Australians set up a commission to study the matter. Last week, the commission made its report. The embryos should be thawed, the panel said, because donation of embryos to someone else would require the consent of the "producers," and no such consent had been given. The panel also held that the embryos in their present state had no life or rights and thus could be destroyed.

The commission members were conscious of treading on slippery legal and ethical grounds. Therefore, they urged that three months be allowed for public opinion to respond to the commission recommendation. Should there be an overwhelming outcry against destroying the embryos, the commission would reconsider.

Couples now enrolling in Sydney's Queen Victoria hospital for in vitro fertilization programs must specify what should be done with the embryos if something happens to them.

This assures that a situation similar to the Rioses won't recur. But what of other complex questions? In France, a woman recently had to go to court to be allowed to bear a child from her deceased husband's frozen sperm. How should such a request be handled? What should be done if a surrogate mother breaks her child-bearing contract and refuses to give up the infant she had promised to bear for someone else?

Our society has failed so far to come up with enforceable rules for curbing the destructive potential of atomic power. We are reaping the nightmarish harvest for that failure. The possibilities of misuse of scientists' ability to advance or retard procreation are manifold. Ethical and legal boundaries need to be set before we stray too far.

In the English version, respondents are instructed to underline the sentence that explains what the Australians did to help decide what to do with the embryos. Note that the question does not ask what they ultimately did with the embryos, but rather focuses on the process used to make the decision. The fifth paragraph begins with the sentence that provides the answer: "The Australians set up a commission to study the matter."

BEST COPY AVAILABLE

Canada

Question 6. Soulignez la phrase où l'on explique ce qu'ont fait les Australiens pour aider à décider ce qu'il fallait faire des embryons congelés appartenant à un couple tué dans un écrasement d'avion.

EDITORIAL

La technique suscite le besoin de nouvelles règles

La SCIENCE a le don de devancer les lois et l'éthique. L'explosion de la bombe atomique en 1945 l'a démontré de façon spectaculaire au point de vue de la destruction de la vie, et les techniques pour vaincre l'infertilité humaine le démontrent aujourd'hui sur le plan de la création de la vie.

La plupart d'entre nous se sont réjouis avec la famille anglaise Brown lorsque le premier bébé-éprouvette, Louise, a vu le jour. Nous nous sommes également émerveillés devant d'autres grandes premières — tout récemment, la naissance de bébés en santé qui étaient auparavant des embryons congelés qu'on attendait d'implanter au moment propice dans l'utérus de la future mère.

C'est d'ailleurs deux embryons congelés qui ont soulevé une tempête de questions juridiques et éthiques en Australie. Ces embryons devaient être implantés dans l'utérus d'Elsa Rios, épouse de Mario Rios. Une implantation précédente avait échoué et les Rios voulaient essayer encore une fois de devenir parents. Ils ont cependant péri dans un écrasement d'avion avant de pouvoir le faire.

Qu'est-ce que l'hôpital australien allait faire des embryons congelés? Pouvait-il les implanter dans l'utérus d'une autre femme? Plusieurs se sont porté volontaires. Les embryons avaient-ils un droit quelconque sur l'importante succession des Rios? Ou devaient-ils être détruits? Les Rios n'avaient rien prévu pour l'avenir des embryons, cela se comprend.

Les Australiens ont créé un comité pour étudier la question. La semaine dernière, ce comité a publié son rapport. Il a conclu que les embryons devaient être décongelés, car leur donation à une autre personne aurait exigé le consentement des «reproducteurs», lequel n'avait pas été donné. Le comité a aussi conclu que les embryons, dans leur état actuel, n'avaient ni vie ni droits et qu'ils pouvaient donc être détruits.

Les membres du comité étaient conscients d'être sur un terrain glissant au point de vue légal et éthique. Par conséquent, ils ont exigé un délai de trois mois pour permettre au public de réagir à leur recommandation. Si celui-ci s'opposait fortement à la destruction des embryons, le comité aurait alors réexaminé la question.

Les couples qui s'inscrivent au programme de fécondation in vitro de l'hôpital Queen Victoria de Sydney doivent désormais préciser ce qui doit être fait des embryons si quelque chose leur arrive.

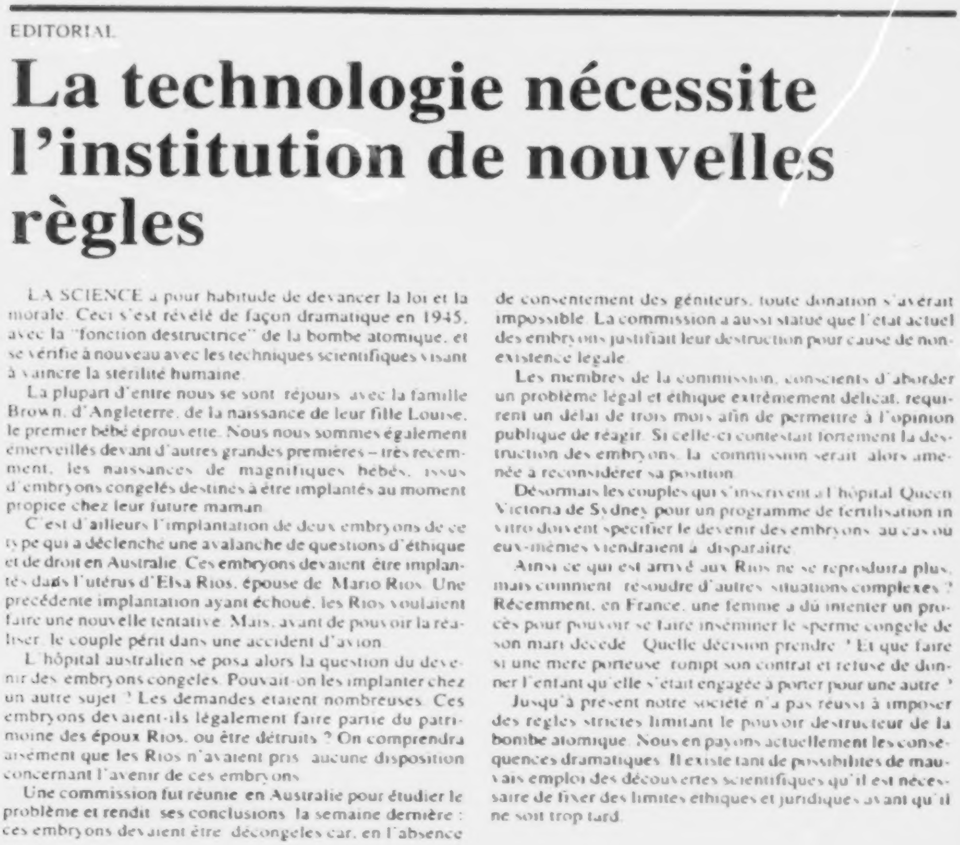
Cela permet d'éviter qu'une situation semblable à celle des Rios ne se reproduise. Mais qu'en est-il d'autres questions complexes? En France, une femme a récemment dû s'adresser à un tribunal pour pouvoir se faire inséminer par le sperme congelé de son défunt mari. Comment devrait-on répondre à une telle demande? Et que devrait-on faire si une mère porteuse ne respecte pas son contrat de grossesse et refuse de remettre l'enfant qu'elle s'est engagée à porter pour quelqu'un d'autre?

Jusqu'à maintenant, notre société n'a pas réussi à établir des règles exécutoires pour limiter le pouvoir destructeur de la bombe atomique. Nous subissons aujourd'hui les conséquences désastreuses de cet échec. Les possibilités d'utiliser la science à mauvais escient pour hâter ou retarder la procréation sont multiples. Il est nécessaire d'établir des limites éthiques et juridiques avant que nous nous écartions trop du droit chemin.

BEST COPY AVAILABLE

France

Question 6. Soulignez la phrase indiquant ce que les Australiens ont fait pour décider du sort des embryons congelés appartenant à un couple mort dans un accident d'avion.



EDITORIAL

La technologie nécessite l'institution de nouvelles règles

LA SCIENCE a pour habitude de devancer la loi et la morale. Ceci s'est révélé de façon dramatique en 1945, avec la "fonction destructrice" de la bombe atomique, et se vérifie à nouveau avec les techniques scientifiques visant à vaincre la stérilité humaine.

La plupart d'entre nous se sont réjouis avec la famille Brown, d'Angleterre, de la naissance de leur fille Louise, le premier bébé éprouvette. Nous nous sommes également émerveillés devant d'autres grandes premières – très récemment, les naissances de magnifiques bébés, issus d'embryons congelés destinés à être implantés au moment propice chez leur future maman.

C'est d'ailleurs l'implantation de deux embryons de ce type qui a déclenché une avalanche de questions d'éthique et de droit en Australie. Ces embryons devaient être implantés dans l'utérus d'Elsa Rios, épouse de Mario Rios. Une précédente implantation ayant échoué, les Rios voulaient faire une nouvelle tentative. Mais, avant de pouvoir la réaliser, le couple périt dans une accident d'avion.

L'hôpital australien se posa alors la question du devenir des embryons congelés. Pouvait-on les implanter chez un autre sujet ? Les demandes étaient nombreuses. Ces embryons devaient-ils légalement faire partie du patrimoine des époux Rios, ou être détruits ? On comprendra aisément que les Rios n'avaient pris aucune disposition concernant l'avenir de ces embryons.

Une commission fut réunie en Australie pour étudier le problème et rendit ses conclusions la semaine dernière : ces embryons devaient être décongelés car, en l'absence de consentement des géniteurs, toute donation s'avérait impossible. La commission a aussi statué que l'état actuel des embryons justifiait leur destruction pour cause de non-existence légale.

Les membres de la commission, conscients d'aborder un problème légal et éthique extrêmement délicat, requirent un délai de trois mois afin de permettre à l'opinion publique de réagir. Si celle-ci contestait fortement la destruction des embryons, la commission serait alors amenée à reconsidérer sa position.

Désormais les couples qui s'inscrivent à l'hôpital Queen Victoria de Sydney pour un programme de fertilisation in vitro doivent spécifier le devenir des embryons au cas où eux-mêmes viendraient à disparaître.

Ainsi ce qui est arrivé aux Rios ne se reproduira plus, mais comment résoudre d'autres situations complexes ? Récemment, en France, une femme a dû intenter un procès pour pouvoir se faire inséminer le sperme congelé de son mari décédé. Quelle décision prendre ? Et que faire si une mère porteuse rompt son contrat et refuse de donner l'enfant qu'elle s'était engagée à porter pour une autre ?

Jusqu'à présent notre société n'a pas réussi à imposer des règles strictes limitant le pouvoir destructeur de la bombe atomique. Nous en payons actuellement les conséquences dramatiques. Il existe tant de possibilités de mauvais emploi des découvertes scientifiques qu'il est nécessaire de fixer des limites éthiques et juridiques avant qu'il ne soit trop tard.

Two aspects of the text could affect the difficulty of this task. The most obvious difference between the versions is that the Canadian version uses a four-column format. Because there are fewer words per line within the column, it is easier to skim for the correct answer; fewer eye fixations per line, and therefore less eye movement, are needed.

A second difference has to do with the punctuation used in the target sentence. In the Canadian French version, the target response is found in a simple sentence that covers just two lines. In France's version, the target response is included in a much longer sentence which covers more than four lines and is divided by a colon. More material than the required information is included in this longer, more complex sentence. Consequently, a French respondent would not only have to locate the particular sentence, but would also need to determine how the test makers define a sentence. Is it the whole sentence, from the beginning "Une" to the final period, including more than the required information? Or does the desired sentence end at the colon? These differences are certain to make this item more difficult for the French to answer correctly.

BEST COPY AVAILABLE

9.5 Do the Examples Fit the Theory?

To this point, we have examined variations among the different versions of particular assessment items and have predicted how these variations would affect item difficulty. The question now is, how can we know if the differences had the predicted effects? What item statistics would help us make this determination? The simplest statistic, the p-value, serves as a good point of departure. This statistic tells us what percentage of the population within each country answered the item correctly. This statistic is insufficient in and of itself, however. If there were marked differences in the overall literacy abilities of the various countries, we would expect the p-values for any particular item to reflect these differences and vary accordingly.

To overcome this limitation, researchers calculated the average p-value for each block of items for each country. This information allows us to compare the direction and magnitude of differences between item and block p-values to determine how item difficulty might differ across item versions. Table 9.1 provides this information.

Table 9.1: **Item and block p-values by language and country**

	English		French		
Item/block	United States	Canada	Canada	Switzerland	France
Example 1	.69	.68	.60	.47	.43
Block	*.71*	*.71*	*.62*	*.68*	*.57*
Example 2	.70	.80	.83	.70	.40
Block	*.64*	*.69*	*.60*	*.68*	*.52*
Example 3	.62	.64	.63	.80	.59
Example 4	.78	.81	.70	.88	.76
Block	*.74*	*.75*	*.71*	*.78*	*.60*
Example 5	.55	.50	.34	.35	.37
Example 6	.59	.56	.40	.66	.57
Block	*.67*	*.66*	*.58*	*.66*	*.54*
Example 7	.58	.65	.49	.52	.20
Block	*.71*	*.71*	*.62*	*.68*	*.57*

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The item p-values as well as the block p-values vary across each row. This suggests that there are differences in overall literacy proficiencies among the nations.

For any one of the example items, we can compare the item p-value and the block p-value in a particular country to see whether the item was more or less difficult for this population than were the other items in the block. Consider Example 5. In each country, the item p-value is lower than the block p-value. We can therefore conclude that in all of the countries this item is, on average, more difficult than others in the same block. In contrast, consider Example 2. The p-value for this item is higher than the corresponding block p-value in every country except France. For the French, this item would be biased. In other words, when individuals in France and individuals in other countries with the same level of literacy proficiency are given this item, it is harder for the French than for those in the other countries to answer the item correctly.

Next let us compare the magnitude of the differences between item and block p-values for the United States and Canada, and for the three French-speaking samples. Although identical test materials were used in the United States and Canada, there are some variations between the two countries in the differences between item and block p-values. The differences among the French versions tend to be larger, however.

Extending this analysis, it is helpful to compare the magnitude and the direction of the differences between item and block p-values across the countries. These data, presented in Table 9.2, highlight the direction of the effect of changes in item difficulty.

Table 9.2: **Differences between item and block p-values by language and country**

	English		French		
Example	United States	Canada	Canada	Switzerland	France
1	-.02	-.03	-.02	-.21	-.14
2	+.06	+.11	+.28	+.02	-.12
3	-.12	-.13	-.08	+.02	-.01
4	+.04	+.06	-.01	+.10	+.16
5	-.12	-.16	-.24	-.31	-.17
6	-.08	-.10	-.18	0	+.03
7	-.13	-.06	-.13	-.16	-.37

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Based on these statistics, the underlying model would suggest that:

- Example 1 would be hardest for the Swiss French, somewhat hard for the French, and easiest for the Americans and the Canadians.
- Example 2 would be relatively difficult for the French and easiest for the French Canadians.
- Example 3 would be relatively easy for the Swiss French and French.
- Example 4 would be easiest for the French.
- Example 5 would be hardest for the Swiss French.
- Example 6 would be hardest for the French Canadians.
- Example 7 would be hardest for the French.

Do the observed item statistics match those suggested based on our evaluation of differences among the item texts? To some extent they do. Specifically, they match for four of the seven examples. Analysis of the items does not provide insights as to why French Canadians find example 2 relatively easy, however, or why the Swiss French have so much difficulty with example 5.

9.6 Conclusions

Although the methods used in this analysis are rather crude, the results suggest that the framework underlying the IALS item development process is consistent across two languages and five cultures. In other words, the cognitive processes underlying literacy appear to be universal. The framework does not account for all the differences across countries, but the amount of variance explained by the task and material characteristic variables is substantial, as demonstrated by Kirsch and Mosenthal. The remaining differences across countries are therefore likely to be quite small.

In sum, this study did not yield answers to all questions concerning why certain items did not behave the same across countries. Nevertheless, the work represents a potentially important breakthrough and marks a methodological improvement over past international studies. In particular, this work underscores the importance of paying very close attention to the specific characteristics of text and tasks in developing equivalent assessments across countries. Without a doubt, more rigorous study is needed. Conducting a controlled experiment within each country and language would be extremely valuable, as would having independent groups of raters make the predictions about item difficulty.

Chapter 10:

Scaling and Scale Linking

Kentaro Yamamoto

10.1 Introduction

Two major goals of the International Adult Literacy Survey were to report literacy proficiency results for many different countries and different languages on common scales, and to have these scales correspond to the three scales that were used to report the results of the National Adult Literacy Survey (NALS), conducted in the United States in 1992. This chapter of the report describes the models and procedures used to scale the International Adult Literacy Survey results and to link the IALS results to the NALS scales.

In the IALS, the objective was to compare the performance of groups of individuals who were administered different assessment forms. As noted earlier in this report, a variant of matrix sampling was used in the survey, so that different respondents received different sets of items. Although the different assessment forms were designed to measure the same cognitive skills domain, they may differ, unintentionally or by design, in their level and range of difficulty. Accordingly, in reporting the survey results, it is inappropriate to use statistics based on the number of correct responses, such as the proportion of items answered correctly. Differences in total scores (or statistics based on them) between respondents who took different sets of items may be caused by differences in respondents' abilities or by differences in difficulty between the two sets of items, or both. Moreover, item-by-item reporting does not provide any information about the distribution of skills within various subpopulations of interest.

To make valid comparisons of performance in an assessment in which participants receive different sets of items, all results need to be expressed on a common scale. One way to construct a common scale is to directly or indirectly translate results from each test form into a number-correct scale of a particular base or reference form. In this method, results from one test form (say, Form Y) are made comparable to results from a reference test form (say, Form X) in some specifically defined sense. Data collection designs and statistical procedures for establishing such comparability are often referred to as *test equating procedures* (see Petersen, Kolen, and Hoover 1989). In equipercentile observed-score equating, for example, number-correct scores on Form Y are transformed to have the same distribution as number-correct scores on Form X for a specified population (Braun and Holland 1982). Alternative procedures equate "true-scores" (i.e., expected test scores) on test forms X and Y (e.g., Levine 1955).

Another way to overcome the limitations of conventional scoring methods is to use item response theory (IRT) scaling. IRT scaling methods make it possible to summarize the performance of a sample of respondents on a series of common scales even if respondents are given different assessment items.

When several items require similar skills, the response patterns for these items should have some regularity. This regularity can be used to characterize both respondents and items in terms of a common standard scale, even when all respondents do not take the same set of items. A common scale makes it possible to analyze distributions of performance in a population or subpopulation and to estimate the relationships between proficiency and background variables. As in the earlier NALS, three literacy scales were used to report the IALS results: prose, document, and quantitative.

Using common scales also makes it possible to determine the relative difficulty of various literacy tasks in the assessment. Each IALS task was assigned a proficiency score based on an analysis of response probabilities. Specifically, a task's score (or "difficulty value") is the point on the scale at which respondents have a high probability of performing the task successfully. For the IALS, a response probability of 80 percent was the criterion used, so the corresponding scale score values are referred to as "RP-80s." What does this mean? If, for example, a prose task has an RP-80 of 278, this means that adults with a prose score of 278 have about an 80 percent probability of answering the task correctly. Individuals with higher scores have an even higher probability of success on the task, while those with lower scores have lower probabilities of success. Tables presenting the RP-80 values for the IALS items are included in Appendix K. These tables show that the difficulty of the assessment tasks varied widely.

It is also important to point out that regardless of what procedure is used to aggregate data, a certain amount of information is regarded as non-essential to the analysis. The methods and procedures employed to analyze the IALS results were carefully designed to capture most of the dominant characteristics of the data.

10.2 Analysis of Items Using Mantel-Haenszel Statistics

Holland (1985) proposed the use of the Mantel-Haenszel (MH) procedure as a practical way to identify test items that function differently in two groups of examinees. The MH statistics are used to examine the adequacy of applying common item parameters to multiple populations.

The standard MH procedure was applied to examine items used in the IALS analyses. The notational scheme and terminology used here follow the convention described in Holland and Thayer (1988). One group is referred to a *focal* group, F, and the performance of this group of examinees is of primary interest. The other group is referred to as a *reference* group, R, and the performance of this group is taken as a standard. Notations such as A_j, B_j, C_j, and D_j represent the numbers of respondents who belong to a corresponding group and who produced a response given an equivalent ability (matching two groups) level j.

		Response		
		Correct	Incorrect	Row total
Group	Reference	A_j	B_j	n_{Rj}
	Focal	C_j	D_j	n_{Fj}
Column total		m_{1j}	m_{0j}	T_j

MH-CHISQ is given as

$$MH-CHISQ = \frac{\left(\left|\sum_j A_j - \sum_j E(A_j)\right| - \frac{1}{2}\right)^2}{\sum_j Var(A_j)}$$

where $E(A_j)$, and $Var(A_j)$ are defined as

$$E(A_j) = \frac{n_{Rj} \cdot m_{1j}}{T_j}$$

$$Var(A_j) = \frac{n_{Rj} \cdot n_{Fj} \quad m_{1j} \cdot m_{0j}}{T_j^2 \cdot (T_j - 1)}$$

Under H_0: A/B=C/D, MH-CHISQ has an approximate chi-square distribution with 1 degree of freedom, (χ^2=7.9, p=0.005), (χ^2=6.6, p=0.01), (χ^2=5.0, p=0.025), and (χ^2=3.8, p=0.05).

The most lenient notion of comparability was used for these analyses; that is, all participants were included without weights. The two groups were matched using only the total scale scores within a booklet. Since each item appeared in at least three of the seven assessment booklets, the MH statistics were calculated three times. Tables presenting the average MH statistics by item and by country are included in Appendix K. Dropped items and countries with unique item parameters are not included in the tables but were included in calculating the statistics. Data for the United States were not included, however, because of the different BIB design employed there.

All items considered here were constrained to have the same item parameters in at least seven of the countries. These analyses identify what appear to be several statistically significant differences. These differences cannot, however, be interpreted literally. MH statistics are known to be over-estimated under two conditions: first, where two groups have different ability distributions, and second, where the discrimination parameter (the slope of the item response function) is large. An additional factor which inflates MH scores is the use of a single matching criteria, i.e., total score correct within a booklet. This criteria disregards differences in the rate of completion of the assessment among countries. Thus, the MH statistic will be greatly inflated where either none, or a substantial proportion, of respondents quit in the middle of the assessment. A high total non-response rate will not, however, have a significant impact on the MH statistic. For example, Canada/English had the least increase of not-reached responses (4%) and Swiss/German had the highest increase

(17%) and both had the highest average MH statistics among all nine countries compared. Given the amount of the inflation of MH statistics associated with these increased not-reached proportions, the application of asymptotic distributions of Chi squared statistics should be performed with caution. Moreover, the comparison of nine countries is being performed simultaneously, thus a Bonferroni adjustment for the MH statistic criteria might be more appropriate. With the adjustment of multiple comparisons, i.e., p*(number of comparisons=9) the MH statistic which is greater than 10.8 is significant at p=.009. In order to take these sources of MH inflation into account, researchers at Educational Testing Service examined plots of observed curves along with the MH statistics and conclude that the IALS test functioned in a reliable way across all tested populations.

10.3 Block Order Effects

Block order effects are a common concern in the field of psychometrics, and participants at a meeting on IALS analysis methodology voiced concerns about this issue. The booklet design used in the survey, referred to as a "balanced-incomplete-block" design, was devised to counterbalance the potential impact of item order on item difficulty. Each block of items appears exactly three times in the set of seven booklets, once in each of the three possible positions: first, second, and last. Each block is also paired with the remaining six blocks in balanced pairs.

The analysis of variance method was used to evaluate the possible block order effect for each IALS country. None of the F statistics was found to be significant, so there is no evidence of a block order effect. Table 10.1 shows the average proportion of items answered correctly, by block, across the IALS countries. The mean proportion corrects were calculated for each country using only those items which were used to estimate proficiency. The "total" column reveals that the average percentage correct was essentially the same for each block position: the values were 76, 78, and 77 for blocks in the first, second, and third positions, respectively.

Table 10.1: Average proportion correct, by block order

Block order	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Total
First	72	78	74	72	77	83	77	76
Second	78	76	77	78	72	83	79	78
Last	74	77	77	77	77	83	72	77

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

In calculating the weighted proportion correct for an item, the denominator was the total number of attempts. Not-reached responses—that is, consecutively missing responses at the end of a block—were not included in calculating proportion correct.

$$P = \frac{\sum \mathit{correct}}{\sum \mathit{correct} + \sum \mathit{wrong} + \sum \mathit{omit}}$$

Tables 10.2 through 10.7 present additional block-level information, including the number and percentage of items answered correctly, average r-biserial correlations and alpha reliabilities, and average percentage of omitted items and items not reached, by block and by country.

Table 10.2: Average number correct, by block and by country

Block no.	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
Core	4.99	4.78	4.66	5.56	5.36	4.46	5.50	5.17	5.02	4.54
1	8.72	7.68	6.01	9.82	9.41	5.38	9.59	8.73	8.24	7.59
2	8.98	7.25	6.43	9.39	9.78	5.73	9.90	8.27	7.70	7.64
3	8.65	6.99	5.70	9.20	9.37	5.05	10.44	8.69	8.20	6.68
4	11.26	8.64	8.64	12.24	12.96	7.47	11.67	10.83	9.73	9.14
5	9.27	6.71	6.83	10.14	10.49	5.46	11.66	9.36	8.67	7.97
6	8.47	6.87	5.80	9.32	9.29	5.14	10.34	8.74	8.53	6.87
7	8.60	7.88	6.69	9.21	9.78	5.17	10.45	8.52	7.97	7.45

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994

Table 10.3: Average percent correct*, by block and by country

Block no.	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
Core	91	88	83	94	91	79	94	94	95	93
1	72	64	55	71	66	54	73	69	70	71
2	71	62	57	69	70	59	76	68	69	71
3	69	60	52	68	67	52	75	68	69	64
4	76	67	65	75	76	62	74	72	71	74
5	66	58	54	66	66	50	74	66	64	67
6	75	71	60	78	74	59	83	78	78	74
7	68	70	59	68	68	51	74	68	68	70

* Percent correct includes all items regardless of IRT scaling results.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 10.4: Average r-biserial correlation by block and by country

Block no.	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
Core	86	69	66	50	43	63	76	74	99	86
1	59	67	56	57	54	53	59	48	51	61
2	66	63	63	50	55	64	51	49	44	61
3	57	55	58	50	61	55	57	50	51	61
4	67	68	63	53	64	61	62	51	60	60
5	63	62	65	52	58	58	61	50	44	63
6	66	51	59	54	53	55	53	50	51	65
7	69	71	64	57	63	62	67	55	50	56

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 10.5: Average alpha reliability, by block and by country

Block no.	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
Core	90	88	82	68	67	80	78	92	94	95
1	93	93	92	88	84	92	91	90	92	95
2	92	92	92	86	85	94	91	91	92	94
3	92	91	91	86	88	92	88	90	92	94
4	95	96	94	90	90	95	93	93	94	96
5	93	93	93	89	87	92	88	91	90	95
6	92	93	91	87	82	92	82	90	91	95
7	93	95	92	88	85	92	89	91	91	94

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 10.6: Average proportion omitted items, by block and by country

Block no.	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
Core	2	3	3	2	2	11	4	2	3	3
1	2	6	8	3	4	14	9	3	5	5
2	2	7	7	4	4	15	7	5	5	7
3	2	7	8	4	4	13	7	4	4	7
4	2	8	7	4	4	12	10	4	5	4
5	2	5	6	5	3	12	8	3	4	6
6	2	4	6	3	3	10	5	2	4	6
7	4	5	7	4	5	19	11	4	6	7

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 10.7: Average proportion not reached items, by block and by country

Block no.	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
Core	9	10	6	1	2	7	2	9	11	18
1	19	20	27	8	5	33	13	15	22	28
2	16	22	24	9	6	35	13	18	24	28
3	16	22	27	9	7	36	7	15	21	31
4	18	28	26	9	5	33	12	16	22	32
5	17	33	26	10	7	36	8	18	22	30
6	13	26	26	8	4	34	4	13	16	28
7	16	25	24	9	4	32	6	17	21	29

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

10.4 Scaling Model

The scaling model used in the IALS is the two-parameter logistic (2PL) model from item response theory (Birnbaum 1968; Lord 1980). This is a mathematical model for the probability that a particular person will respond correctly to a particular item from a single domain of items. This probability is given as a function of a parameter characterizing the proficiency of that person, and two parameters characterizing the properties of that item. Specifically, the 2PL model gives the probability that a person with unobserved proficiency θ_j answers item i correctly as:

$$P(x_{ij} = 1 | \theta_j, a_i, b_i) = \frac{1}{1.0 + \exp(-Da_i(\theta_j - b_i))} \qquad (1)$$

where θ_j is the proficiency of person j; x_{ij} is the response of person j to item i (1 if correct, 0 if incorrect); a_i is the slope parameter of item i, characterizing its sensitivity to proficiency; and b_i is the location parameter of item i, characterizing its difficulty. Note that this is a monotone increasing function with respect to θ. That is, the conditional probability of a correct response increases as the value of θ increases.

A linear indeterminacy exists with respect to the values of q_j, a_i, and b_i for a scale defined under the two-parameter model. In other words, for an arbitrary linear transformation of q, say $q^* = Mq+X$, the corresponding transformations $a^*_i = a_i/M$ and $b^*_i = Mb_i + X$ give:

$$P(x_{ij} = 1 | \theta^*_j, a^*_i, b^*_i) = P(x_{ij} = 1 | \theta_j, a_i, b_i) \qquad (2)$$

This scale indeterminacy was resolved by setting an origin and unit size of θ to the reported scale means and standard deviations of the young adult literacy assessment, as discussed later in this chapter.

The main assumption of IRT is conditional independence. Conditional independence means that respondents with identical ability have a similar probability of responding correctly to an item regardless of their country membership. This assumption applies to those subsamples in a country who received different sets of items.

In other words, item response probabilities depend only on θ (a measure of proficiency) and the specified item parameters, and not on any demographic characteristics of examinees, or on any other items presented together in a test, or on the survey administration conditions. This enables us to formulate the following joint probability of a particular response pattern x across a set of n items.

$$P(\underset{\sim}{x}|\theta,\underset{\sim}{a},\underset{\sim}{b}) = \prod_{i=1}^{n} P_i(\theta)^{x_i}(1-P_i(\theta))^{1-x_i}$$

Replacing the hypothetical response pattern with the real scored data, the above function can be viewed as a likelihood function that is to be maximized with a given set of item parameters. These item parameters were treated as known for the subsequent analyses.

Testing the assumptions of the IRT model, especially the assumption of conditional independence, is a critical part of the data analyses. Serious violation of the conditional independence assumption would undermine the accuracy and integrity of the survey results.

The IRT model also assumes unidimensionality; that is, performance on a set of items can be accounted for by a single unidimensional variable. Although this assumption may be too strong, the use of the model is motivated by the need to summarize overall performance parsimoniously within a single domain.

10.5 Item Parameter Estimation

Identical item calibration procedures, described here in detail, were carried out separately for each of the three literacy scales. Using a modified version of Mislevy and Bock's (1982) BILOG computer program, the two-parameter logistic IRT model was fit to each item using sample weights. BILOG procedures are based on an extension of the marginal-maximum-likelihood approach described by Bock and Aitkin (1981).

In the equation, $P(\underset{\sim}{x}_{j,g}|\theta,\beta)$ is the conditional probability of observing a response vector $\underset{\sim}{x}_{jg}$ of person j from group g, given proficiency θ and vector of item parameters $\beta = (a_1,b_1,...a_i,b_i)$ and $f_g(\theta)$ is a population density for θ in group g. Prior distributions on item parameters can be specified and used to obtain Bayes-model estimates of these parameters (Mislevy 1984). The proficiency densities can be assumed known and held fixed during item parameter estimation or estimated concurrently with item parameters.

The f_g in the above equation are approximated by multinomial distributions over a finite number of "quadrature" points, where X_k, for $k=1,..,q$, denotes the set of points and $A_g(X_k)$ are the multinomial probabilities at the corresponding points that approximate $f_g(q)$ at $q=X_k$. If the data are from a single population with an assumed normal distribution, Gauss-Hermite quadrature procedures provide an "optimal" set of points and weights to best approximate the integral for a broad class of smooth functions. For more general f or for data from multiple populations with known densities, other sets of points (e.g., equally spaced points) can be substituted and the values of $A_g(X_k)$ may be chosen to be the normalized density at point X_k (i.e., $A_g(X_k) = f_g(X_k)/\Sigma_k f_g(X_k)$).

Maximization of L(ß) is carried out by an application of an EM algorithm (Dempster, Laird, and Rubin 1977). When population densities are assumed known and held constant during estimation, the algorithm proceeds as follows. In the E-step, provisional estimates of item parameters and the assumed multinomial probabilities are used to estimate "expected sample sizes" at each quadrature point for each group (denoted $\hat{N}_{gk}$), as well as over all groups (denoted $\hat{N}_k=\Sigma_g \hat{N}_{g,k}$). These same provisional estimates are also used to estimate an "expected frequency" of correct responses at each quadrature point for each group (denoted $\hat{r}_{gik}$), and over all groups (denoted $\hat{r} = \Sigma_g \hat{N}_g \hat{r}_{gik}$). In the M-step, improved estimates of the item parameters are obtained by treating the $\hat{N}_{g,k}$ and $\hat{r}_{gik}$ as known and carrying out maximum-likelihood logistic regression analysis to estimate the item parameters ß, subject to any constraints associated with prior distributions specified for ß.

When proficiency densities are estimated concurrently with item parameters, the $\hat{N}_{g,k}$ values from the previous E-step are used to update the item parameter estimates as well as estimates of the $A_g(X_k)$'s in the M-step. Both the nonparametric solution and the normal solution for the f_g presented in Mislevy (1984) are implemented in the IALS version of BILOG. Updated estimates of both item parameters and population distributions are used to generate the $\hat{N}_{g,k}$ values for the next E-step which, in turn, are used in the next M-step. The process can be continued until all parameter estimates become stable, providing maximum likelihood estimates of both item and population parameters. For the IALS, however, the EM cycles were continued only until item parameter estimates become stable to a specified degree of precision, since alternative procedures were used to obtain estimates of proficiency distributions (see Chapter 11). The EM cycles were terminated when changes to item parameters were less than .005.

To obtain stable item parameter estimates for the IALS and to simplify scale linking procedures, the data accumulated from previous surveys were included in the calibration sample. In effect, this method of parameter calibration puts all survey results on a single provisional common scale. Only linear indeterminacy needed to be resolved in order to align the provisional scale to the reporting scale.

Using the NALS item parameter estimates without modification requires that these items display good fit. Graphical methods and χ^2 statistics were used to verify fit, and the statistics indicated very good fit. Of 76 items, only one item had $\chi^2 = 12.8$, while the balance of the items had χ^2 of less than 9.0. On average, the maximum deviation of observed proportion correct from predicted proportion correct at quadrature points within one standard deviation of the mean was less than 0.005. This is a remarkably small deviation, even more noteworthy because these item parameters were not estimated using the United States IALS data. The fit of NALS item parameters with the United States IALS data supports the use of the original item parameters without modification.

Sample weights were used during item calibration. It is known that different subpopulation distributions occur within different assessment samples. Such variations may arise because of differences in the characteristics of the target populations, the sampling design, or the randomness of sampling. For example, oversampling racial/ethnic minority populations is often necessary to ensure accuracy in estimating group proficiencies. In such cases, the unweighted sample would not correctly represent the targeted population. Post-stratified weights take into account the sampling design including the effects of over-sampling, as well as the randomness of the real data. By applying post-stratified weights, vital characteristics of the sample can be closely matched to the characteristics of the population.

During calibration, the fit of item parameters is maximized about the proficiency distribution of the calibration sample. When item parameters are being estimated, it is ideal to match the proficiency distribution of the calibration sample as closely as possible to that of the population. It is even more critical when item calibration is done on the combined proficiency distribution of multiple assessment samples with great differences in proficiency distributions, such as the IALS.

To obtain unbiased parameter estimates, proficiency distributions for the separate assessment samples were estimated during calibration. Each country received a separate empirical prior normal proficiency distribution according to the response probability, distributions that were updated every iteration. It is known that the samples for each assessment came from somewhat different populations with different characteristics. The calibration procedure should take into account the possibility of systematic interaction of samples and items to estimate unbiased estimates of sample distributions and item parameters. For that reason, a normal distribution with a unique mean and variance for the population of each country was estimated concurrently with item parameters. The reason for not using a multinomial distribution for the prior distribution is the inherent instability of the shape of the distribution due to over-parameterization.

Estimated item parameters for the items on each literacy scale, by country, are presented in the appendices.

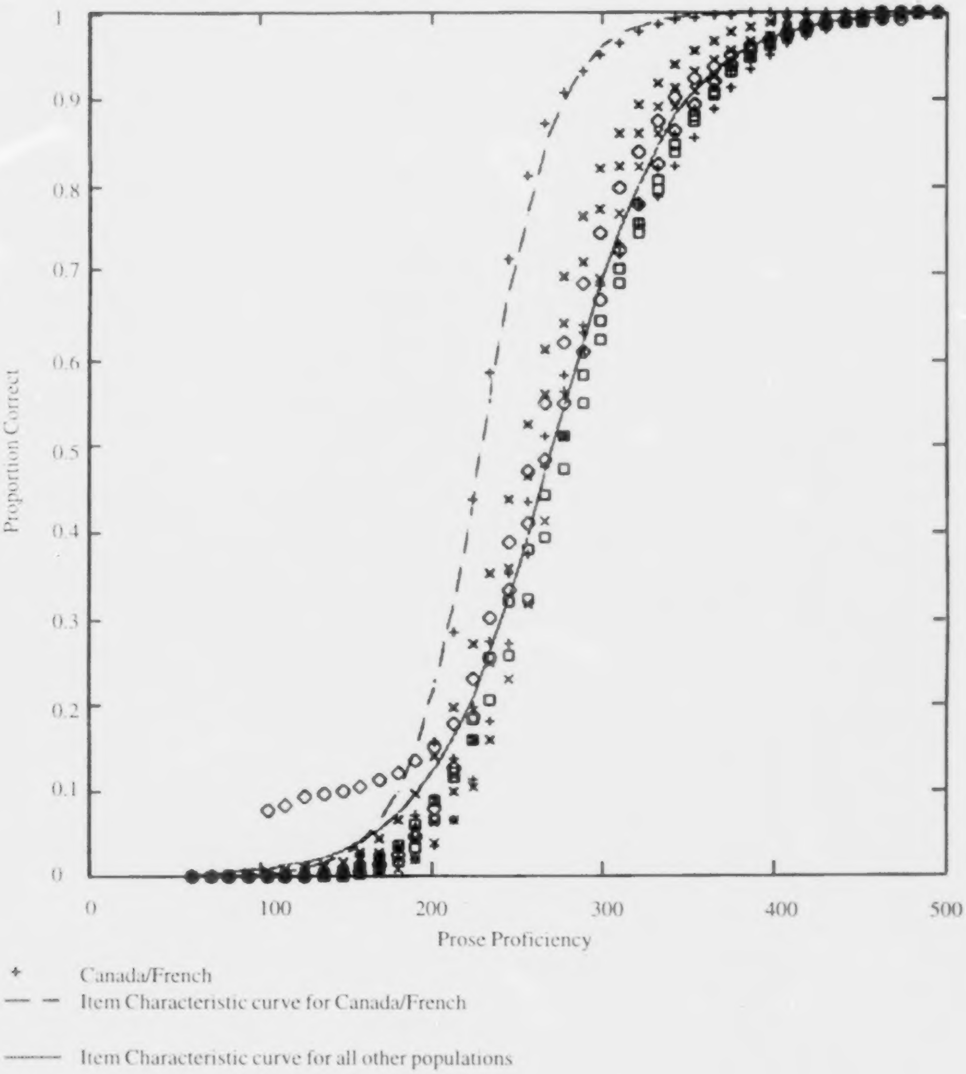
Model fit was evaluated at the item level by examining likelihood ratio chi-square statistics for each country's sample. The fit was also evaluated by inspecting residuals from fitted item response curves. In general, the fit of the model to the data was quite good. The great majority of responses were accurately described by the common item parameters. For some items, there was evidence that the estimated parameters did not fit as well to the assessment samples from a few countries as compared to the other countries. This pattern did not occur consistently for any one country sample, however.

The plot shown in Figure 10.1, for example, clearly indicates that the Canada/French data do not support the use of common item parameters for this particular item.

In this plot, the smooth lines are the fitted two-parameter logistic item response curves. The legends are the (approximate) expected proportions of correct response at various points along the provisional proficiency scale for each subpopulation. The horizontal axis represents the prose scale. Most of the population is distributed within the 200 to 350 score range. This plot indicates that the observed proportions are correct, given that proficiencies are quite similar among most of the countries. The Canada/French data for this item differ noticeably from the common item characteristic curve represented by the solid line. Thus, a unique set of item parameters was estimated for that sample, resulting in two smooth curves for this item.

There are two options for accommodating the misfit of the IRT model while keeping the international scales intact. The quickest approach is to drop the deviant items from the analysis. A drawback of this option is that it results in a smaller number of items, especially if items are dropped when the IRT functions differ in only one or two countries. We would use this approach if the IRT model did not fit at all—for example, if the response function was negative, or if all observed response functions were so far apart from each other that one set of item parameters would not describe responses from most of the countries.

Figure 10.1: **Item response curve for an IALS prose item**



When more than three countries' response data for a particular item showed poor fit to the item parameters common to the rest of the countries, the item was dropped from the analyses. In all, 12 such items were dropped from the calibration. An additional quantitative item was dropped due to colinearity with the item that followed it. (Scores on these two items were virtually identical, which violated the assumption of local independence, so one item was dropped from the analysis.) When data from one to three countries fit poorly with the common item parameters, unique item parameters were estimated only for countries displaying poor fit for the item. In some cases, poor fit could be explained by errors or ambiguities in the printed materials, but in other cases, no obvious reasons could be identified.

Table 10.8 presents the number of common items by country by scale. The assumption of common item parameters for all countries means that there will be 1,010 constraints (that is, 114 items minus 13 dropped items equals 101 items, times 10 country/language samples). Of these 1,010 cases, unique item parameters were required in 81 cases (or 8 percent). In other words, 92 percent of the constraints support a common scale for all 10 samples.

Table 10.8: **Items common to country samples, by literacy scale**

Scale	Dropped	Number of country samples in common: 7	8	9	All 10	Total no. of items
Prose	5	2	7	10	15	39
Document	7	4	5	9	16	41
Quantitative	1	0	4	12	17	34
Total	**13**	**6**	**16**	**31**	**48**	**114**

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Item parameter treatments (common, unique, or dropped) for each item and each IALS country are presented in Appendix K.

Researchers evaluated the extent to which having a subset of item parameters differ among countries would have an impact on test characteristic functions. Three figures show that the deviation between observed and theoretic proportions correct are less than 0.01 for the ability range of 150 to 400, where nearly 99 percent of respondents perform. One exception is found in the extreme end of the document scale. Even the largest deviation for the document scale is extremely small and has almost no impact in terms of defining the proficiency distribution for subpopulations. For example, the lowest and the highest levels do not have a constant interval limit; that is, the ranges include proficiency values up to 225 for level 1, and above 325 for levels 4 and 5. This is strong evidence that systematic bias does not exist for any of the data from all countries over the entire range of ability.

10.6 Scale Linking Procedures

When assessment results are to be reported using an item response theory (IRT) proficiency metric, as is the case for the IALS, establishing a common metric is synonymous with expressing IRT item parameter estimates on a common scale. Procedures for obtaining item parameter estimates on a common scale are referred to here as *scale linking procedures*. Some of the standard approaches to scale linking are briefly summarized in Appendix K. Other published sources provide more detailed information (e.g., Vale 1986; Petersen, Kolen, and Hoover 1989).

As noted earlier, one purpose of the IALS was to develop three literacy scales that could be used to report proficiency results across multiple languages. The other was to relate these three scales to the NALS scales. Linking the IALS results to the NALS scales would allow researchers and policy makers to make inferences from past analyses conducted using the NALS database, the most extensive database on adult literacy in English.

The IALS was based on the same construct of literacy abilities and included some of the same assessment items as the NALS (see Table 10.9). The U.S. IALS sample served as a linking sample between the IALS and NALS scales. In all, 190 items were administered in the U.S. IALS, 76 of which had been administered in the NALS. The remaining 114 were new items developed for the IALS.

Table 10.9: **Composition of U.S. IALS item pool**

Literacy scale	Number of items common to NALS	Number of IALS items	Total
Prose	25	39	61
Document	25	41	66
Quantitative	26	34	60
Total	**76**	**114**	**190**

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Because the IALS item set was largely new, and therefore different from that in the NALS, there was no guarantee that the two assessments were truly measuring the same skills. Further, the number of items common to the IALS and NALS was insufficient to provide accurate linkages between the two assessments. An explicit linking procedure was therefore required to secure a minimal error of measurement.

As noted earlier, the U.S. sample for the main IALS assessment served as a linking sample between the IALS scales and the NALS scales. Every respondent in the U.S. IALS sample received two of the seven new IALS blocks as well as one of the seven NALS blocks. (Respondents in other countries received three of the seven new IALS blocks.) Table 10.10 provides information about the booklet pairings for the linking study. Blocks A through G are IALS blocks, while Blocks 1 through 7 are NALS blocks.

Table 10.10: Booklet pairing for the IALS linking study

	Block letter/number													
B#	A	B	C	D	E	F	G	1	2	3	4	5	6	7
1	2	3						1						
2		2	3						1					
3			2	3						1				
4				2	3						1			
5					2	3						1		
6						2	3						1	
7	3						2							1
8	1		3								2			
9		1		3								2		
10			1		3								2	
11				1		3								2
12					1		3	2						
13	3					1			2					
14		3					1			2				
15	1			2									3	
16		1			2									3
17			1			2		3						
18				1			2		3					
19	2				1					3				
20		2				1					3			
21			2				1					3		

Note: Blocks A through G are IALS blocks. Blocks 1 through 7 are NALS blocks.

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The IALS booklet design made it possible to link the IALS and NALS scales at the item level—one of the most powerful linking designs. As explained earlier in this chapter, the two parameter logistic IRT model was used to calibrate item parameters, and the multiple imputation method (see Chapter 11) was used to incorporate background information as well as to reduce measurement error. Researchers validated the NALS item parameters on the United States sample to ensure the feasibility of keeping the item parameters the same. Fit statistics and plots of item operating characteristics supported the use of identical item parameters from the NALS for the IALS, without modifications.

The least-square procedure by Stocking and Lord could have been applied to the IALS, but the method has some problems. First, it cannot detect whether the new items were on a different scale. Second, it disregards possible differences in ability distributions between the NALS population and the combined international populations. Although such differences may not be critical where there is a large number of items of varying difficulty, the average proportion correct for the IALS was about 0.70, indicating that it was easier than some achievement tests.

The simultaneous estimation methods used for the IALS linking offer advantages in these two respects, but also present disadvantages in terms of scaling complexity. Further, any correction of item responses results in the need to rescale.

After the NALS item parameters were validated, the IALS item parameters were estimated using the total IALS sample together with the United States sample. The means and standard deviations for the population distribution of each country were estimated simultaneously and used to set the prior distributions for the subsequent step. Use of single prior distribution for every country would result in systematic bias toward the prior distribution, with the amount of bias corresponding to the distance of the true distribution from the prior.

The shape of the population distribution was constrained as normal during the item parameter estimation process. Without this constraint, the model would have been unidentified; that is, more than one combination of multinomial proficiency distributions and item parameters might fit nearly as well to describe the data. Multiple imputation (discussed in Chapter 11) is able to correctly accommodate any deviation from this constraint, however. The scaling procedure was repeated for all three scales.

Once the plausible values were on the provisional scale, the next step was to transform them to the NALS scale for comparison. Since the IALS scales were implicitly linked during the item parameter calibration, the accuracy of the NALS scales is the only issue of concern at this stage. The IALS scales adapted the transformation constants directly from the NALS scales.

It is helpful to summarize briefly how the NALS scales were established. Some of the items administered in the 1985 Young Adult Literacy Survey (YALS) were the same as those included in the NAEP 1984 reading assessment. Relying on the common items from the two assessments, the YALS sample proficiency distribution was placed on the NAEP reading scale, a 0 to 500 metric. The mean and standard deviation of the plausible values for the YALS samples were estimated to be 296.6 and 49.0, respectively. The mean and standard deviation of the other three scales—prose, document and quantitative—were also set to these values.

In the NALS, item parameters from the YALS were re-estimated using a larger sample and more accurate procedures than were available at the time of the 1985 analysis. These new item parameters are best suited for comparing distributions of performance for different samples. However, the new sets of item parameters on the provisional scales and the old transformation constants used to produce the 1985 YALS scales would not necessarily produce identical results for the 1985 YALS sample. Thus, new linear transformation constants for the YALS sample were found to match the mean and standard deviation of current plausible value distribution of the YALS sample based on the new item parameters. The same constants were applied to the NALS sample proficiency distribution. The transformation that was applied is as follows: $\theta=A\theta^*+B$ where θ^* is the provisional scale from item calibration and θ is the reported scale.

Table 10.11 presents the transformation constants and the mean and standard deviations for the distributions of the three scales. These constants were derived from the IALS/NALS linkup study and applied to each international provisional scale.

Table 10.11: **Transformation constants applied to provisional scale to produce the reported scale**

Literacy scale	α	β
Prose	51.67	269.16
Document	52.46	237.50
Quantitative	54.41	276.87

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

10.7 Evaluation of Differential Group Performance

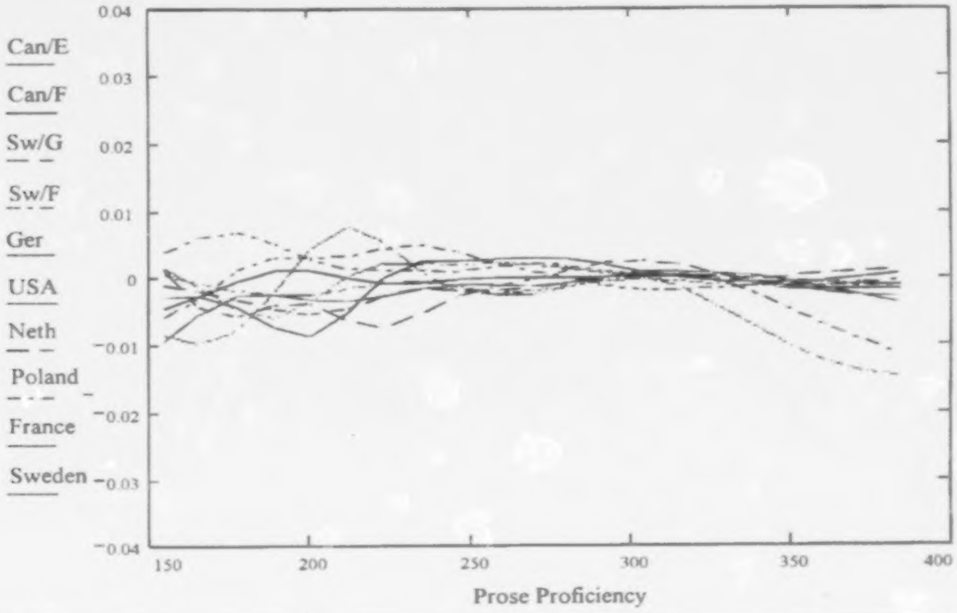
Differences in performance across subpopulations were examined by constructing empirical characteristic curves of tests rather than of items, for major subpopulations defined by variables such as sex and ethnicity.

Yamamoto and Muraki (1991) have found that sets of estimated item parameters, each estimated on separate calibration samples with different racial/ethnic compositions, differed significantly even after an appropriate linear transformation was applied to account for scale indeterminacy. This suggests some degree of differential item functioning (DIF) by racial/ethnic subpopulations. The test as a whole functioned equivalently, however, suggesting that the effect of different item parameters on the subpopulation proficiency estimates would be negligible. In fact, after a linear scale transformation to account for the scale indeterminacy was applied to the data, the estimates of subgroup proficiency distributions, using a different set of item parameters, were virtually identical. Since our concern is with the presence of systematic bias against a particular subpopulation, it is more appropriate to evaluate differential group performance at the test level than at the item level. Therefore, the deviation of the empirical test characteristic curves from the theoretical test characteristic curves was estimated for the IALS countries.

The following figures illustrate the difference between the average empirical proportion correct and the expected proportion correct for the items in a scale for each country's sample. Each point on the scale was estimated in two steps. For every item, the empirical proportions correct were calculated for each sample whose proficiency values were in the selected range. The percents correct were then averaged for all items in the scale. The theoretical proportion correct was calculated based on the estimated item parameters and averaged for all items in the scale. This procedure was repeated for each subpopulation of interest.

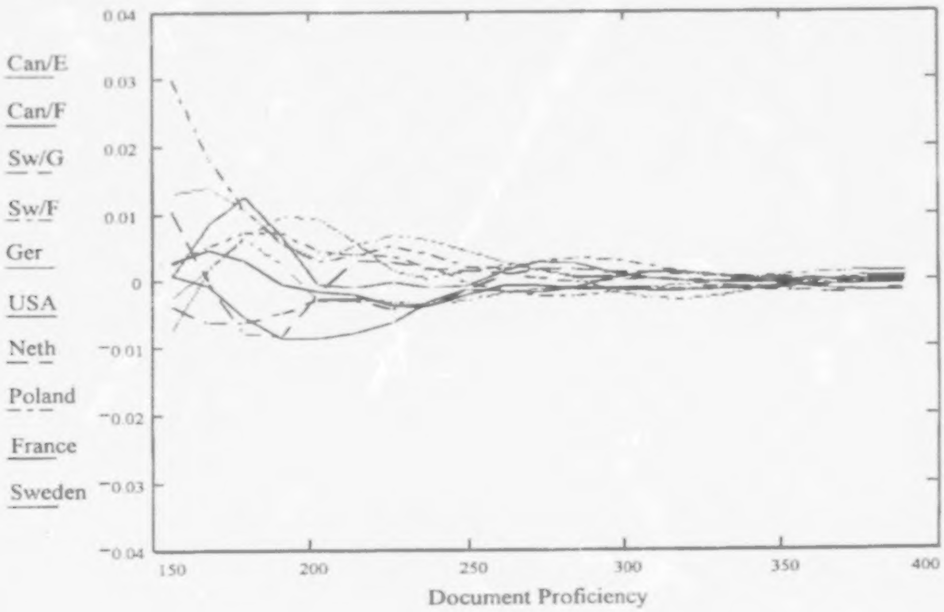
Although some figures show deviations in the test characteristic curves at the low (below 200) end of the proficiency range, the number of individuals performing in this range is very small, and therefore stable estimates cannot be made. Test characteristic curves should be compared only in the part of the proficiency range where most of the population scores.

Figure 10.2: **Prose scale test characteristics by country (observed-theoretic) proportion correct**



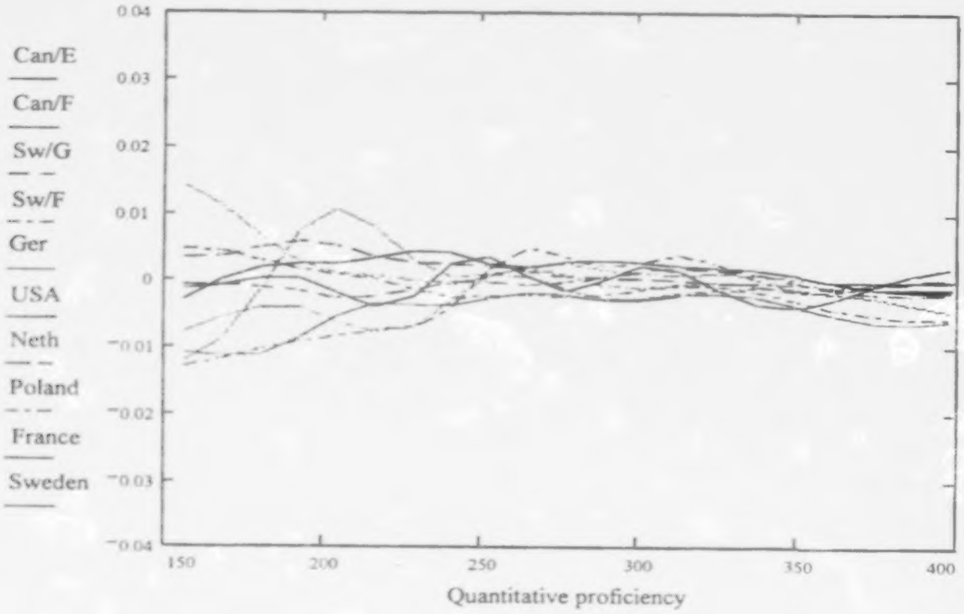
Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Figure 10.3: **Document scale test characteristics by country (observed-theoretic) proportion correct**



Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Figure 10.4: **Quantitative scale test characteristics by country (observed-theoretic) proportion correct**



Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

If the test characteristic curves deviated systematically within a subpopulation of interest, this could be viewed as evidence that the test was functioning differentially (in other words, was biased) for that group. The absolute values of deviation for the 200 to 350 range on the scales are less than 0.01 and practically 0.

10.8 Conclusions

As in other applications of IRT, item parameter estimates must be used in the IALS because the item parameters are unknown. Research is underway on how uncertainty associated with item parameter estimates affects the estimation of proficiency (see, for example, Tsutakawa and Johnson 1990). The estimation error associated with scale linking constants represents another source of uncertainty in the results of IRT-based assessments. Some preliminary investigations into estimating the uncertainty associated with scale linking constants has been carried out by Sheehan and Mislevy (1988) and Johnson, Mislevy, and Zwick (1990). At the present time, the standard errors for the IALS estimates reflect only error due to respondent sampling and imputation (that is, measurement error).

Scale linking in the IALS was a complicated process, involving the use of common items. Concurrent calibration approaches were followed. The linking procedures used in the IALS are evolving in response to the changing needs of the survey and other factors, and research into the adequacy of various methods continues.

As with all statistical estimation procedures, the scale-linking process is subject to errors. When scale linking involves the concatenation of several linking steps, such errors carry over to the next scale, particularly when a sequential linking procedure is used. To reduce the possibility of propagating such linking errors, researchers conducting the IALS analyses attempted to avoid scale linking beyond item calibration.

Psychometric analysis of the IALS data—such as linking the IALS scales to the U.S. National Adult Literacy Survey (NALS) scales and estimating item response parameters, discussed in this chapter—yields considerable evidence of the validity, reliability and precision of the assessment at both the item level and the assessment level. Although they did not evaluate the evidence in detail, the experts believe that IALS was successful in achieving psychometric equivalence in the assessment.

Chapter 11:

Proficiency Estimation

Kentaro Yamamoto and Irwin S. Kirsch

11.1 Introduction

Most cognitive skills assessments are concerned with accurately assessing the performance of individual respondents for the purposes of diagnosis, selection, or placement. Regardless of which measurement model is being used, classical test theory or item response theory, the accuracy of these measurements can be improved—that is, the amount of measurement error can be reduced—by increasing the number of items given to the individual. Thus, achievement tests containing more than 70 items are common.

When analyzing the distribution of proficiencies in a group of persons, however, more efficient estimates can be obtained from a sampling design similar to that used in the IALS. Such designs solicit relatively few cognitive responses from each sampled respondent but maintain a wide range of content representation when responses are summed for all respondents. The advantage of estimating population characteristics more efficiently is offset by the inability to make precise statements about individuals.

In the IALS, plausible values methodology was used to estimate respondents' literacy proficiencies based on their performance on the literacy tasks. This methodology is described in the first part of this chapter. Special procedures used to impute the literacy proficiencies of individuals who did not complete the cognitive assessment are discussed in the latter part of the chapter.

11.2 Plausible Values and Proficiency Estimation

The multiple imputation procedure for proficiency estimation used in the IALS is based on the plausible values methodology developed by Mislevy (1991), following the work by Rubin (1987). This methodology was developed as a way to estimate key population features consistently and to approximate others at least as well as standard IRT procedures would. Although this approach does not provide the best estimates of individual scores, it is well suited for population estimates. The uncertainty of individual estimates is retained in order to increase the accuracy of the population estimates. Specifically, unbiased proficiency estimates for subpopulations are obtained by incorporating the relationships between individual background variables and the posterior proficiency distribution. The multiple values are sampled from the posterior distribution directly instead of summarized into a point estimate.

It cannot be emphasized too strongly that plausible values are not test scores for individuals in the usual sense. In other words, they are unlike the more familiar ability estimates of educational measurement which are, in some sense, optimal for each respondent (e.g., maximum likelihood estimates and Bayes estimates). Point estimates that are optimal for individual respondents can produce decidedly nonoptimal (inconsistent) estimates of population characteristics (Little and Rubin 1983). Plausible values are constructed explicitly to provide consistent estimates of population effects, even though they are not generally unbiased estimates of the proficiencies of the individuals with whom they are associated (Mislevy, Beaton, Kaplan, and Sheehan 1992).

A detailed review of plausible values methodology is given in Mislevy (1991). Along with theoretical justifications, Mislevy presents comparisons with standard procedures, discusses biases that arise in some secondary analyses, and offers numerical examples. The following description summarizes key aspects of the plausible values approach, focusing on its implementation in the 1995 IALS analyses.

11.3 Proficiency Estimation in the IALS

Let y represent the responses of all sampled respondents to background questions, and let θ represent the scale proficiency values. If θ were known for all sampled examinees, it would be possible to compute a statistic $t(\theta,y)$—such as a scale or composite subpopulation sample mean, a sample percentile point, or a sample regression coefficient—to estimate a corresponding population quantity T.

Because the scaling models are latent variable models, however, θ values are not observed even for sampled respondents. To overcome this problem, the IALS analysis follows Rubin (1987) by considering θ as missing data, and approximates $t(\theta,y)$ by its expectation given (x,y), the data that actually were observed, as follows:

$$\begin{aligned} t^{*}(\underline{x},\underline{y}) &= E\left[t(\underline{\theta},\underline{y})\mid\underline{x},\underline{y}\right] \\ &= \int t\left(\underline{\theta},\underline{y}\right)p\left(\underline{\theta}\mid\underline{x},\underline{y}\right)d\theta \qquad (1) \end{aligned}$$

It is possible to approximate t^{*} using random draws from the conditional distribution of the scale proficiencies given the item responses x_j, background variables y_j, and model parameter for sampled respondent j. These values are referred to as imputations in the sampling literature, and as plausible values in the IALS.

The value of θ for any respondent that would enter into the computation of t is thus replaced by a randomly selected value from his or her conditional distribution. Rubin (1987) proposed to repeat this process several times so that the uncertainty associated with imputation can be quantified by "multiple imputations." For example, the average of multiple estimates of t, each computed from a different set of plausible values, is a numerical approximation of t^{*} of the above equation. The variance among them reflects uncertainty due to not observing θ. It should be noted that this variance does not include the variability of sampling from the population.

Plausible values for each respondent j are drawn from the conditional distribution $P(\underline{\theta}_j|x_j,y_j,G,S)$, where G is a matrix of regression coefficients and S is a common variance matrix for residuals. Using standard rules of probability, the conditional probability of proficiency can be represented as follows:

$$P(\underline{\theta}_j|x_j,y_j,G,S) \propto P(x_j|\underline{\theta}_j,y_j,G,S)P(\underline{\theta}_j|y_j,G,S)$$

$$= P(x_j|\underline{\theta}_j)P(\underline{\theta}_j|y_j,G,S) \qquad (2)$$

where θ_j is a vector of three scale values, $P(x_j|\underline{\theta}_j)$ is the product over the scales of the independent likelihoods induced by responses to items within each scale, and $P(\underline{\theta}_j|y_j,G,S)$ is the multivariate joint density of proficiencies of the scales, conditional on the observed value y_j of background responses and parameters G and S. Item parameter estimates are fixed and regarded as population values in the computation described in this section.

In the IALS analyses, a normal multivariate distribution was assumed for $P(\underline{\theta}_j|y_j,G,S)$, with a common variance, S, and with a mean given by a linear model with slope parameters, G, based on the first approximately principal components of several hundred selected main effects and two-way interactions of the complete vector of background variables. The background variables included sex, ethnicity, language of interview, respondent education, parental education, occupation, and reading practices, among others. Based on the principal component method, components representing 99 percent of the variance present in the data were selected. These included principal components, referred to as conditioning variables, are denoted as y^c. The following model was fit to the data.

$$\theta = \Gamma y^{c} + \varepsilon$$

where ε is normally distributed with mean zero and variance S. As in a regression analysis, G is a matrix, each column of which is the effects for one scale and S is the three-by-three matrix variance of residuals between scales.

Note that in order to be strictly correct for all functions G of θ, it is necessary that $p(\theta|y)$ be correctly specified for all background variables in the survey. In the NALS, principal component scores based on nearly all of the background variables were used. Those selected variables were chosen because of their policy relevance. The computation of marginal means and percentile points of θ for these variables is nearly optimal. Estimates of functions T involving background variables not conditioned on in this manner are subject to estimation error due to misspecification (Mislevy 1991). The magnitudes of these errors diminish as each respondent provides more cognitive data—that is, responds to a greater number of items. The magnitude of these errors appears to be negligible in the IALS (i.e., biases in regression coefficients below 5 percent) as compared to the NALS, due to the larger numbers of cognitive items presented to each IALS respondent (on average, 16 items per scale).

IALS respondents who could not or did not proceed beyond the background questions represent an exception to this rule. These individuals did not attempt the assessment tasks because they were unable to read or write English, or because of a physical or mental disability, or because they simply refused to participate in the survey. Excluding these respondents from the survey would have severely distorted the proficiency estimates for some of the IALS populations and subpopulations. The procedures used to estimate the proficiencies of these non-responding individuals are discussed later in this chapter.

The basic method for estimating G and S with the EM procedure for a single scale was described in Mislevy (1985). The EM algorithm requires the computation of the mean, θ and variance, S, of the posterior distribution in (2). For the multiple scales of IALS, the computer program CGROUP (Thomas 1993) was used. The program implemented a method to compute the moments using higher order asymptotic corrections to a normal approximation. Case weights were employed in this step.

After completing the EM algorithm, the plausible values were drawn in a three-step process from the joint distribution of the values of G for all sampled respondents who attempted five or more cognitive items. First, a value of G was drawn from a normal approximation to $P(G,S|x_j,y_j)$ that fixed S at the value $\hat{S}$ (Thomas 1993). Second, conditional on the generated value of G (and the fixed value of $S=\hat{S}$), the mean q, and variance S_j^p of the posterior distribution in the equation (2) were computed using the same methods applied in the EM algorithm. In the third step, the q were drawn independently from a multivariate normal distribution with mean q and variance S_j^p. These three steps were repeated five times, producing five imputations of q for each sampled respondent.

For those with an insufficient number of responses, the G and Ss described in the previous paragraph were fixed. Hence, all respondents—regardless of the number of items attempted—were assigned a set of plausible values for the three scales.

The plausible values could then be employed to evaluate equation 1 for an arbitrary function T as follows:

1) Using the first vector of plausible values for each respondent, evaluate T as if the plausible values were the true values of θ. Denote the result T_1.
2) In the same manner as in step 1, evaluate the sampling variance of T, or $Var(T_1)$, with respect to respondents' first vectors of plausible values. Denote the result Var_1.
3) Carry out steps 1 and 2 for the second through fifth vectors of plausible values, thus obtaining T_u and Var_u for u=2, . . .,5.
4) The best estimate of T obtainable from the plausible values is the average of the five values obtained from the different sets of plausible values:

$$T. = \frac{\sum_u T_u}{5}$$

5) An estimate of the variance of T. is the sum of two components: an estimate of $Var(T_u)$ obtained as in step 4 and the variance among the T_us:

$$Var(T.) = \frac{\sum_u Var_u}{5} + (1+\frac{1}{5})\frac{\sum_u (T_u - T.)^2}{5-1}$$

The first component in *Var(T.)* reflects uncertainty due to sampling respondents from the population. The second component reflects uncertainty due to the fact that sampled respondents' θs are not known precisely, but only indirectly through x and y.

11.4 Classical Test Statistics and Proficiency Values

Proportion of correct responses

In addition to examining the average literacy proficiencies of IALS respondents in each country, it is also useful to study classical test statistics such as the proportion of correct responses. Final weights were used to calculate mean plausible values and average proportion correct on each country's cognitive items for respondents aged 16 to 65. Proportions correct included all items, even those dropped from the IRT-based analysis.

It is well known that for a fixed set of items, mean proficiency values and proportion correct statistics have a strong linear relationship. As Table 11.1 shows, such a relationship was observed in the IALS data for all three scales. The order and the relative differences among the country/language subpopulations are essentially the same with proficiency values and with proportion correct.

Table 11.1: Mean proficiency values and proportion correct

	Canada/ English	Canada/ French	France	Germany	Nether- lands	Poland	Sweden	Switzer- land/ French	Switzer- land/ German	United States	Mean	r
Prose												
Proficiency	284	264	233	276	283	229	301	265	263	274	267	97
% correct	70	64	52	65	66	52	73	64	65	69	64	
Document												
Proficiency	284	265	239	285	287	224	306	274	270	268	270	95
% correct	74	69	62	77	75	60	80	75	76	74	72	
Quantitative												
Proficiency	286	266	245	293	288	235	306	280	280	275	275	97
% correct	67	60	58	69	68	54	73	69	67	66	65	

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Tables presenting the proportion correct statistics by item and by country are included in Appendix L.

Non-response

The observed proportion of non-response is sometimes used as an indicator of sampled individuals' motivation to participate in the assessment, but this is not necessarily appropriate because, as noted earlier in this report, there are different types of non-response.

In the IALS, omitted responses were defined as any missing responses that occurred before a valid response. Not-reached responses, on the other hand, were defined as consecutively missing responses at the end of a block of items. These two types of non-response were treated differently in the IALS data analyses, as discussed below. Omitted responses, which have more direct consequences for proficiency estimation, are discussed first. Not-reached responses, discussed second, have no direct impact on proficiency estimation.

Omitted responses

Omitted responses were scored as wrong. This decision was based on the interpretation that respondents had an opportunity to examine and attempt to answer items positioned before an item with a valid response, and the lack of response indicates that they were unable to respond correctly. It is known, however, that less able respondents are more likely than able respondents to leave cognitive items unattempted.

Let us consider the possibility that the proportions of omitted responses in the data indicate the motivation of respondents within a country/language sample and contribute to differences in the proficiency distributions. Table 11.2 shows the proficiency means and average proportion of omitted responses by scale. The correlations of proficiency value means and average omitted proportions for the three scales were calculated and have 8 degrees of freedom for the 10 country/language populations. None of the correlations was found to be significantly different from 0, even at the p=.10 level, where the critical value is .622 with df=8.

This clearly indicates that proficiency scores are not related to the number of omitted responses, and the proportion of omitted responses cannot be used to predict the relative performance of any population. (This fact becomes even more apparent when plots are examined.) The proportions of omitted responses were highest in Poland and Sweden. In fact, excluding the Polish data points, the correlations for the prose, document, and quantitative scales are 0.1, 0.0, and 0.1, respectively. This highlights the irrelevance of omitted responses to the subpopulation proficiency means.

Table 11.2: **Mean proficiency values and proportion of omitted responses**

	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States	Mean	r
Prose												
Proficiency	284	264	233	276	283	229	301	265	263	274	267	-.37
% omitted	2.04	4.35	5.80	4.12	3.69	10.04	8.51	3.94	4.15	3.52	5.02	
Document												
Proficiency	284	265	239	285	287	224	306	274	270	268	270	-.53
% omitted	1.49	4.21	4.17	1.95	2.64	7.79	5.32	2.40	2.65	1.73	3.43	
Quantitative												
Proficiency	286	266	245	293	288	235	306	280	280	275	275	-.40
% omitted	1.39	4.06	4.59	2.52	3.05	7.80	6.89	2.15	3.02	2.48	3.80	

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The weighted percentages of omitted responses by item and by country are presented in the appendices.

It is often observed that more difficult items are more likely to have omitted responses. This relationship can be evaluated by examining the correlations between the proportion of items correct for each scale, an indicator of difficulty, and the proportion of omitted responses. If the occurrence of omitted responses is due to item difficulty, then the correlations between the proportions of correct responses and omitted responses should have significantly negative values. Table 11.3 shows these results. Degrees of freedom for the three scales are 35, 37 and 30, respectively.

The critical values for the scales, given the degrees of freedom, are .282, .282, and .306 for p=0.1, and .467, .467, and .471 for p=0.01.

Table 11.3: **Correlations between proportions of correct and omitted responses by country**

	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States	Mean
Prose	-.62	-.70	-.74	-.64	-.66	-.79	-.73	-.65	-.64	-.42	-.66
Document	-.41	-.49	-.51	-.45	-.32	-.42	-.59	-.48	-.31	-.39	-.44
Quantitative	-.40	-.34	-.54	-.42	-.65	-.55	-.70	-.36	-.57	-.42	-.48

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Of the 30 correlations examined (i.e., three literacy scales times 10 country/language populations), 18 were significant at the 0.01 level, 11 were significant at the 0.05 level, and 1 was significant at the 0.1 level. These data clearly show that omitted responses are influenced by item difficulty. Sweden, France, and Poland had the highest average correlations.

Some confounding factors are involved, however. Specifically, correct responses, incorrect responses, and omitted responses together make up the entire set of valid responses. Thus, omitted responses and incorrect responses combined have an upper bound. To partially cancel the potentially strong relationship between the omitted responses and proportions correct for a country, means of proportion correct across the 10 country/language populations were used as indicators of item difficulty. As presented in the earlier table, the average proportions correct within a scale vary a great deal across the countries due to the proficiency differences observed among them. The three countries with the highest correlation averages also had the largest gaps in the average proportion correct. These data indicate that the bounded omitted responses have little impact on the higher correlations between omitted responses and proportions correct. The results were nearly identical, and the averaged correlations differed less than 0.01 from the results reported in the earlier table.

If lack of motivation can be inferred from an increase in omitted responses, there should be a positive relationship between the percentage of omitted responses and item location. None of the correlations between the proportions of omitted responses and item order was significantly different from zero for every country and every block. The sign of the correlation tended to be negative, however, indicating a tendency toward a decrease in omitted responses with the order of items. It should be emphasized that the IALS assessment items were not ordered in terms of difficulty within each block. Further, responses to the last item in each block were excluded from the calculation of correlations because, by definition, it is not possible to have omitted responses at the end of a block.

Not-reached responses

Not-reached responses can be classified into two groups: non-participation immediately or shortly after the background information was collected, and premature withdrawal from the assessment after a few cognitive items were attempted. The first type of not-reached response can vary a great deal across countries according to the frames from which the samples were selected. For example, in the United States, a pre-existing sample pool was used, and most of the sample was accustomed to taking part in a governmental data collection effort. Such experience did not include cognitively demanding tasks of the type included in the IALS, however. This unfamiliarity seems to have led to higher rate of immediate withdrawal from the cognitive assessment among United States IALS participants as compared to NALS participants.

The second type of not-reached response is due to quitting the assessment early, resulting in incomplete cognitive data. Not-reached items were treated as if they provided no information about the respondent's proficiency, so they were not included in the calculation of likelihood functions for individual respondents. Therefore, not-reached responses have no direct impact on the proficiency estimation for subpopulations. The impact of not-reached responses on the proficiency distributions is mediated through the subpopulation weights. The mechanism is analogous to the impact of post-stratified weights on population estimates. With the post-stratified weights, the correct population composition can be represented and performance characteristics can be estimated. The major assumption underlying such estimations is that the missing cognitive data are similar to the observed data *when the background information is the same*. This is especially important in estimating the proficiencies of respondents who immediately quit after the background questions were answered. Their performance distribution was based on the observed responses of respondents with similar characteristics and backgrounds.

By definition, not-reached responses always increase with item order. It is important to compare the differential increases in not-reached responses among language/country populations, however. One may argue that a substantial increase in not-reached responses for a country is due to a lack of motivation across the entire population.

Researchers examined the proportions of not-reached responses at two points in a block: the first item and the last item. The not-reached proportion for the initial item in a block is higher than the standard calculation. Because every block appears in three different booklets in three different positions, blocks in the second and third position have inflated not reached proportions. Accordingly, the averages of not-reached proportions over three different positions across seven blocks were used to examine changes in the not-reached proportion. The results are presented in Table 11.4.

The average proportion of not-reached responses to the first item in a block was highest for Poland, followed by the United States, French Canada, and France; the proportions were lowest for the Netherlands, Germany, and Sweden. The most pronounced increases in the proportion of not-reached responses were observed in data from German Switzerland and Poland.

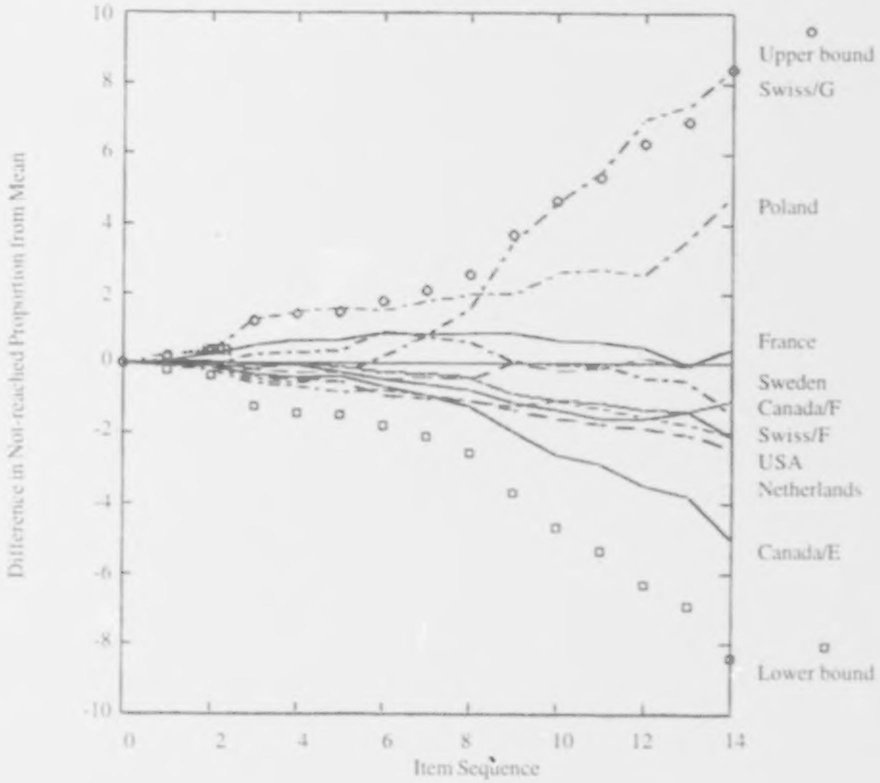
Table 11.4: **Average proportion of not-reached responses by location in the blocks (in percents)**

	Canada/ English	Canada/ French	France	Germany	Nether- lands	Poland	Sweden	Switzer- land/ French	Switzer- land/ German	United States	Mean
First	15	22	22	6	3	29	5	14	15	27	16
Increase	4	8	10	9	7	15	10	7	17	7	9
Last	19	30	32	15	10	44	15	21	32	34	25

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994

The plot in Figure 11.1 shows the variability of the increase in the proportion of not-reached response by country and by item order. Each line represents the difference in the proportions of not-reached responses of a subpopulation from the means of 10 subpopulations. The two legends of box and diamond represent 2.26 times the standard deviation of the statistic at a given ordinal position. The coefficient of 2.26 was selected for a T-distribution with 9 degrees of freedom. If the differences in the not-reached proportion from the mean fall between the range represented by two legends, then the statistics can be considered as not different from zero. Except at the 11th to 13th position of items for German-speaking Switzerland, all statistics are within the desired bounds. The increase in the not-reached responses reveals only small differences among most of the country/language populations.

Figure 11.1: **Difference in not reached proportion from mean by item sequence**



Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994

11.5 Missing Cognitive Data

In any survey of opinions, ability, or demographic characteristics, it is extremely difficult to obtain a 100 percent response to all questions. The most common and least desirable way to treat missing data is to ignore it. This approach is acceptable if one can assume that missing cases are missing at random and that the remaining observed cases are representative of the target population. In this case, the result would be slightly larger standard errors due to reduced sample size, and the other estimates would remain unbiased. Randomly missing data rarely occur in real data collections, however, especially in surveys of performance. If the incidence of non-response varies for major subgroups of interest, or if the missing responses are related to the measurement objective—in this case, literacy—then inferring the missing data from observed patterns results in biased estimates.

Ignoring missing data may be the only option if nothing is known about differences between the respondent population and the missing population. There are several reasons why one might know very little about non-respondents. Some sampled households may be vacant, for example, or the residents may refuse to participate out right. In these cases, the post-stratified weights account for the non-participating subpopulation based on the assumption that the information is missing at random within a subpopulation.

Many large-scale assessment programs have excluded non-responding cases from analyses. Although a proportion of the missing data and some of the characteristics of the missing data sample are sometimes reported, the impact of non-response on the results of these surveys is often not determined. When response rates vary across subpopulations, however, reporting the results based only on the sample of respondents can yield both biased and inaccurate proficiency distributions for some subpopulations.

In conducting analyses for the National Adult Literacy Survey in the United States, researchers at Educational Testing Service implemented an innovative approach to minimize distortions in the population proficiency estimates due to non-response. This approach is described at length in Appendix L. The following section discusses the adaptation of this approach in the IALS.

11.6 Imputing the Literacy Proficiencies of Non-respondents

To minimize bias in the proficiency estimates due to non-response, special procedures were implemented to impute the literacy proficiencies of non-respondents. As noted earlier in this report, for the purposes of this study, non-respondents were defined as those who completed fewer than five literacy tasks—the minimum number needed to estimate their proficiencies directly.

When an individual sampled for the IALS decided to stop the assessment, the interviewer used a standardized non-response coding procedure to record the reason why the person was stopping. If the individual was unable to provide this information, the interviewer asked the family member or friend who was assisting in the interview to indicate the reason.

The stated reasons were subsequently used to classify individuals who stopped the assessment into two groups:

- Individuals who cited *literacy-related reasons* for stopping the assessment—for example, they failed the core task booklet, had language difficulties, had reading or writing difficulties, had a learning disability, had a mental or emotional condition, had a speech impairment, or were mentally retarded; and

- Individuals who cited *reasons unrelated to literacy* for stopping the assessment—for example, they refused to answer the main booklet, partially completed the main booklet, had a hearing impairment, were blind or had a visual impairment, had a physical disability or other disability, or cited other unspecified reasons.

Although there were many reasons for non-completion, most of the cases were explained by a small handful of reasons (Table 11.5). In fact, of the 1,358 individuals who stopped the assessment for literacy-related reasons, the vast majority were persons who either failed the core task booklet (670 persons, or 49 percent) or had language difficulties (522 individuals, or 38 percent). More than 81 percent of the 1,638 adults who stopped the assessment for reasons unrelated to their literacy skills were individuals who refused to participate in the survey.

Most non-respondents who cited language difficulties were foreign born (92 percent), and almost half (43 percent) had less than a secondary education. The other major literacy-related reason, failing the core task booklet, was attributable mostly to the native-born (just 19 percent of those who failed were foreign-born) and to persons with less than a secondary education (56 percent).

On the other hand, just 14 percent of the adults who refused to participate were foreign born; 19 percent of those who refused had less than a secondary education. The foreign born and less educated were slightly overrepresented among non-respondents as compared with respondents.

Table 11.5: **Among non-respondents,* proportions of foreign born and individuals with less than a secondary education, by reason for non-response and by country**

		Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States	Total
Literacy related												
Failed core	N	64	15	154	6	16	338	17	3	4	53	670
Foreign born	%	49	0	22	49	45	4	64	53	0	41	19
Less than secondary ed.	%	71	60	61	22	47	59	37	33	0	22	56
Language difficulty	N	92	36	65	0	12	0	0	79	118	120	522
Foreign born	%	100	100	87	-	100	-	-	90	89	88	92
Less than secondary ed.	%	56	86	42	-	41	-	-	22	19	58	43
Other	N	17	41	27	0	8	7	1	4	0	61	166
Foreign born	%	25	12	13	-	16	0	0	29	-	12	14
Less than secondary ed.	%	35	86	41	-	73	90	0	0	0	50	57
Total	N	173	92	246	6	36	345	18	86	122	234	1,358
Foreign born	%	74	45	38	49	59	4	60	86	86	58	46
Less than secondary ed.	%	59	83	54	22	51	60	35	22	18	48	51
Non-literacy- related												
Refusal	N	153	79	229	46	18	340	7	79	64	318	1,333
Foreign born	%	30	5	14	43	27	3	16	25	19	11	14
Less than secondary ed.	%	19	10	29	11	29	33	26	7	2	8	19
Other	N	28	14	34	2	4	15	39	5	5	159	305
Foreign born	%	60	0	5	49	0	0	66	14	0	7	19
Less than secondary ed.	%	13	36	36	0	65	44	7	0	0	14	18
Total	N	181	93	263	48	22	355	47	84	69	477	1,638
Foreign born	%	35	4	13	44	22	3	57	24	17	10	15
Less than secondary ed.	%	18	14	30	1	29	34	10	6	2	10	19

* That is, individuals who completed fewer than five literacy tasks

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

When individuals cited a literacy-related reason for not completing the cognitive items, this implies that they were unable to respond to the items. On the other hand, citing reasons unrelated to literacy implies nothing about a person's literacy proficiency. Based on these interpretations, IALS adapted a procedure originally developed for the NALS to treat cases in which an individual responded to fewer than five items per literacy scale, as follows:

- If the individual cited a literacy-related reason for not completing the assessment, then all consecutively missing responses at the end of a block of items were treated as wrong.
- If the individual cited reasons unrelated to literacy for not completing the assessment, then all consecutively missing responses at the end of a block were treated as "not reached."

It is important to emphasize that all responses actually made by examinees were kept as they were; that is, they were not overwritten by any other method.

Proficiency values were estimated based on respondents' answers to the background questions and the cognitive items. As an intermediate step, the functional relationship between these two sets of information was calculated, and this function was used to obtain unbiased proficiency estimates with reduced error variance. A respondent's proficiency was calculated from a posterior distribution that was the multiple of two functions: a conditional distribution of proficiency, given responses to the background questions; and a likelihood function of proficiency, given responses to the cognitive items.

The first method used to handle non-responding cases in which an individual stopped the assessment for reasons unrelated to literacy has no impact on the likelihood function. The calculation of a likelihood function is only affected by the method used to handle cases in which non-respondents cited literacy-related reasons for stopping the assessment. The likelihood function for this group will be very peaked to the lower end of the scale, a result which is believed to represent their proficiency accurately. With this procedure, summary statistics can be produced for the entire population including those who respond to cognitive items correctly in various degrees as well as those who did not, either because they would not or because they could not do so.

The procedure described here assumes that adults' self-reported reasons for non-response are accurate and reliable. The accuracy of the information about reasons for non-response is particularly important because of its impact on the proficiency distributions, particularly for some subpopulations. Analyses of background data largely affirmed non-respondents' reasons for stopping the assessment. It is highly unlikely that this level of consistency would have occurred if the reasons for non-response were erroneous.

11.7 Conclusions

Missing responses are a fact of life in surveys of opinions, ability, or demographic characteristics. Although the most common way to treat missing data is to ignore it, this practice assumes that missing cases are missing at random and that the observed samples are representative of the target population. If response rates are different among subpopulations, however, this practice yields both biased and inaccurate proficiency distributions for some subpopulations, and thus for the total population as well.

In the IALS, as in the NALS, information collected from survey participants was used to classify non-respondents into two groups: those who stopped the assessment for literacy-related reasons (for example, individuals with language problems, extreme reading difficulties, or some type of cognitive disability), and those who stopped for reasons unrelated to literacy (for example, adults who simply refused to respond to the assessment tasks, and those with physical disabilities). A special procedure was then used to impute the literacy scores of non-respondents.

The accuracy of adults' self-reported reasons for non-response is important because of the potential impact on the proficiency distributions, particularly for some subpopulations. In future assessments of this kind, it might therefore be advantageous,

Chapter 12:

Plausibility of Proficiency Estimates

Richard Shillington

12.1 Introduction

One way to judge the IALS data is to explore the degree to which the literacy estimates are plausible given variations in the distribution of underlying determinants of literacy, such as age and education. Logic would dictate that, if the observed literacy skill profiles can be predicted with any degree of certainty, the distributions themselves must be reliable. This chapter describes a study in which researchers used a series of logit models to assess the degree to which literacy level, as measured by the populations' distribution into literacy levels, is predictable and correlated with independent variables.

12.2 Methodology

This section describes the approach used to assess the predictability of literacy scores based on demographic factors for each of the IALS populations. The purpose of this analysis is to evaluate the extent to which the literacy levels of the population in the database for each nation are predictable based on demographic characteristics.

Scores from the IALS survey were categorized into levels using the score ranges shown in Table 12.1.

Table 12.1: **Score ranges for IALS literacy levels**

Minimum value	Maximum value	Literacy level
0	226.000	1
226.001	276.000	2
276.001	326.000	3
326.001	376.000	4
376.001	999	5

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

In the statistical analyses reported, estimates were generated for literacy levels 1, 2, 3, and 4/5 (combined).[1]

The task of assessing predictability can be operationalized in more than one way. One option is to use ordinary least squares (OLS) regression techniques to estimate the raw literacy score on a continuous scale and then categorize the score to a level. Alternatively, one can use logistic regressions to predict the number of individuals at each literacy level.

The first stage of the research is to determine whether the raw literacy scores or the levels are to be used as the measurement tool. This choice, in turn, influences the choice of statistical technique. If one uses raw scores, then the prediction question may be operationalized as the prediction of the relationship between *average* literacy score and demographic predictors. Ordinary least squares regression techniques might be expected to perform this task, based on well-established practice. If, on the other hand, one is more interested in understanding the relationship between independent variables and low levels of literacy (for example, levels 1 or 2), or high levels of literacy, then a series of logistic regressions does a superior job of prediction compared with ordinary least squares regressions (Shillington 1996).

For the investigation described here, the focus was on predicting literacy levels, so a logistic regression approach was used.

The summary data presented in Table 12.2 are drawn from the international database of IALS results. Poland and France consistently have the highest proportions of their populations at level 1 and the lowest proportions at level 4/5. Sweden consistently has the highest proportion at level 4/5, but Germany has the lowest proportion of its population at level 1 for two of the three literacy scales. Summary statements about average literacy levels rely on how this is measured. The nation with the highest proportion at level 4/5 does not necessarily have the lowest proportion at level 1, as evidenced by the respective results for Sweden and Germany on the document scale.

---

[1] Many of the IALS analyses have been conducted with levels 4 and 5 combined.

Table 12.2: **Estimated distribution of the populations across the literacy levels, by literacy scale (in percents)**

	Prose literacy levels			
Country	1	2	3	4/5
Canada/English	19	26	31	24
Canada/French	28	26	38	9
France	41	34	22	3
Germany	14	34	38	13
Netherlands	13	31	42	14
Poland	43	35	20	3
Sweden	13	23	36	27
Switzerland/French	18	34	39	10
Switzerland/German	19	36	36	9
United States	21	26	32	21

	Document literacy levels			
Country	1	2	3	4/5
Canada/English	21	24	31	25
Canada/French	31	27	28	14
France	36	32	26	6
Germany	9	33	39	19
Netherlands	13	27	42	18
Poland	45	31	18	6
Sweden	11	22	37	30
Switzerland/French	16	29	39	16
Switzerland/German	18	29	37	16
United States	24	26	31	19

	Quantitative literacy levels			
Country	1	2	3	4/5
Canada/English	20	24	33	23
Canada/French	28	32	30	10
France	33	30	28	9
Germany	7	27	43	24
Netherlands	12	26	43	19
Poland	39	30	24	7
Sweden	11	21	37	31
Switzerland/French	13	24	42	20
Switzerland/German	14	26	41	19
United States	21	25	31	22

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The data indicate that some countries, such as Germany, have low percentages of adults demonstrating level 1 literacy without having a particularly high percentage of adults in level 4/5. Conversely, other nations (such as English Canada) have high proportions of adults in level 4/5 without very low proportions in level 1.

To summarize the population distribution across the literacy levels and to create a summary measure of the amount of variability in literacy levels for each country, the literacy levels were converted to a numerical scale with values of 1, 2, 3, and 4.5, respectively, for levels 1, 2, 3, and 4/5. Averages and standard deviations were then calculated on this scale, and these are presented in Tables 12.3 and 12.4.

The data in Table 12.3 reveal similarities between the literacy values across the three scales for different nations. In other words, nations that do well on the prose scale also tend to do well on the document and quantitative scales.

Table 12.3: **Mean values in literacy levels, by literacy scale**

Country	Prose	Document	Quantitative
Canada/English	2.71	2.72	2.71
Canada/French	2.32	2.32	2.28
France	1.89	2.05	2.19
Germany	2.57	2.78	2.95
Netherlands	2.64	2.75	2.77
Poland	1.85	1.87	2.02
Sweden	2.91	3.01	3.04
Switzerland/French	2.46	2.62	2.80
Switzerland/German	2.39	2.59	2.74
United States	2.64	2.55	2.66

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 12.4: **Standard deviation in literacy levels, by literacy scale**

Country	Prose	Document	Quantitative
Canada/English	1.22	1.24	1.21
Canada/French	1.05	1.17	1.07
France	0.91	1.00	1.08
Germany	1.02	1.03	1.03
Netherlands	1.01	1.07	1.06
Poland	0.89	0.99	1.03
Sweden	1.17	1.16	1.16
Switzerland/French	0.99	1.08	1.09
Switzerland/German	0.98	1.10	1.09
United States	1.20	1.20	1.22

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The countries with the highest literacy levels vary somewhat by scale. In general, however, Sweden, Germany, English Canada, and Switzerland have the highest levels.

The amount of variability in each country's literacy levels also varies, as shown in Table 12.4. For each scale, the variability tends to be somewhat higher for English Canada, the United States, and, to a lesser degree, Sweden. This greater extent of variability may make the prediction of scores somewhat easier in those countries.

12.3 Logistic Regressions

Twelve logistic regressions were fit to the data for each country. The three literacy scales—prose, document, and quantitative—were combined with four regressions, one for each literacy level (1, 2, 3, and 4/5).

The same regression model was estimated for each nation and literacy level. The independent variables included sex, age, age^2, immigration, year of immigration, income quintile, income quintile^2, labor force status, educational attainment, mother's and father's educational attainment, educational attainment^2, and number of hours spent watching television each day. For some countries, the logistic regressions with all variables did not converge to a solution. In these cases, a simplified set of independent variables was used.

A subset of the observations was used for the regressions. This subset excluded observations which had non-response or non-applicable responses for any of the required independent variables. This reduced the sample sizes for the regressions to some extent, as shown in Table 12.5.

Table 12.5: **Observations used in regression before and after removal of non-applicable and non-response observations, by country/language group**

Country	Number used	Number available	Percent used
Canada/English	2,782	3,951	74
Canada/French	1,306	1,709	78
France	2,044	2,996	66
Germany	1,402	2,062	67
Netherlands	2,779	3,090	89
Poland	1,637	3,000	55
Sweden	1,859	3,038	57
Switzerland/French	1,079	1,440	76
Switzerland/German	995	1,398	74
United States	2,070	3,045	70
Total	**17,953**	**25,729**	**69**

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Because of this restriction, Poland and Sweden's sample sizes were reduced by almost half, while the reductions in other countries' samples were more modest. It should be noted that, as a result of this process, the observations used in the logistic regressions may not be strictly representative of the sampled populations. Although it is still possible to compare the analyses for various nations, one must keep in mind that the comparisons are valid as comparisons of the predictability of literacy based on demographic factors for those observations with complete information *on the independent variables*.

12.4 Assessing the Multiple Correlation Between Literacy and Demographic Characteristics

Regression models were used to seek a correlation between predicted and actual levels of literacy for each IALS country. Four logistic regressions were fit to the dependent variables to predict levels 1, 2, 3, or 4/5.

After estimating the regression coefficients, each actual literacy level was compared to the predicted literacy level. The predicted literacy level was defined based on the fitted probabilities resulting from the four logistic regressions.

$$\text{Predicted Literacy}_i = 1 \times P_{i1} + 2 \times P_{i2} + 3 \times P_{i3} + 4.5 \times P_{i4/5}$$

where

P_{ij} is the estimated probability that individual i has literacy in the j^{th} literacy category for i=1..n and j=1, 2, 3, and 4/5.

The analysis using this equation was repeated for all three literacy scales for each of the 10 samples. Table 12.6 shows the correlation between predicted and actual literacy levels by country.

Table 12.6: **Correlation between predicted and actual literacy levels, by literacy scale and by country (in percents)**

Prose		Document		Quantitative	
Country	Correl.	Country	Correl.	Country	Correl.
Germany	42	Germany	40	Germany	37
Sweden	46	Sweden	45	Switzerland/German	40
Switzerland/German	47	Poland	46	Sweden	42
Switzerland/French	49	Switzerland/German	46	Switzerland/French	44
Poland	52	Switzerland/French	50	Poland	47
Netherlands	54	Netherlands	52	Netherlands	50
United States	54	United States	52	United States	56
France	60	Canada/English	55	France	57
Canada/English	61	France	58	Canada/English	59
Canada/French	67	Canada/French	68	Canada/French	72

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The correlations between actual and predicted literacy levels can be interpreted much like the usual correlations from ordinary least squares regression. The pattern in the above correlations is reasonably consistent. On each literacy scale, English and French Canada and France had the highest correlations between predicted and actual literacy levels, while Germany, German Switzerland, and Sweden had the lowest correlations.

A second approach to assessing predictability is to assess the "goodness of fit" of the logistic regressions models. In this context, "goodness of fit" refers to the extent that the population distribution across the literacy levels changes with the prediction variables in a manner predicted by the logistic model.

This is accomplished by arranging the observations in descending order by the predicted probability of being in level 1. The sorted observations are then grouped into deciles of equal size representing "deciles of risk." For each decile, the actual number of individuals in each literacy level is compared to the predicted number. This is repeated for each country and for each literacy scale.

Table 12.7 shows how this approach was employed in assessing the "goodness of fit" of the logistic regression models used to evaluate estimated versus actual prose literacy levels for English-speaking Canada. The top section of the table presents the actual distribution by literacy level in each decile of risk, the middle section presents the predicted distribution, and the bottom section presents the difference between the actual and predicted proportions.

Table 12.7: **Actual and predicted literacy levels for English Canada, prose scale (in percents)**

	Actual			
Decile of risk	Level 1	Level 2	Level 3	Level 4/5
1	76	20	2	1
2	42	36	19	3
3	21	39	32	8
4	11	26	42	21
5	7	21	54	18
6	2	33	49	16
7	3	36	37	24
8	2	12	25	60
9	3	7	30	60
10	0	12	32	56
	Predicted			
Decile of risk	Level 1	Level 2	Level 3	Level 4/5
1	67	23	8	1
2	39	34	23	5
3	22	34	34	9
4	13	31	39	18
5	7	28	42	22
6	5	29	42	24
7	4	24	39	33
8	3	16	31	51
9	1	15	34	50
10	1	14	36	50
	Difference			
Decile of risk	Level 1	Level 2	Level 3	Level 4/5
1	8.6	2.8	5.9	0.1
2	3.5	2.0	3.8	1.7
3	0.8	4.2	2.3	1.1
4	2.0	4.7	3.0	3.7
5	0.2	7.3	12.0	4.4
6	3.4	4.1	6.9	7.6
7	0.5	12.1	1.9	9.7
8	0.5	3.5	5.5	9.5
9	1.6	8.0	3.3	9.7
10	0.6	2.0	3.6	6.2

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The overall error measure is the average of the 40 values in the table which are the differences between the actual and predicted proportions. Stated another way, the decile prediction error is summarized simply as the average of the 40 absolute differences between the observed and predicted proportions at each literacy level. These differences are summed over the literacy levels and over the deciles.

In Table 12.7, the average is the total error term, defined as follows:

$$\begin{aligned}
\text{Error} &= |\,76\% - 67\%\,| + |\,20\%-23\%| + |2\%-8\%| + |1\%-1\%| + \\
&\quad |\,42\% - 39\%| + |\,36\% - 34\%\,| + |\,19\% - 23\%\,| + |\,3\% - 5\%| + \\
&\quad \ldots + |\,56\% - 50\%\,| \\
&= 8.6\% + 2.8\% + 5.9\% + 0.1\% + \\
&\quad 3.5\% + 2.0\% + 3.8\% + 1.7\% + \\
&\quad \ldots \\
&\quad \ldots + 6.2\%
\end{aligned}$$

The quantitative values are not as familiar as correlations, but can still be used to compare countries' literacy levels.

The results in Table 12.8 indicate that the goodness of fit is exceptionally good for several countries. The values are higher but still quite small in English and French Canada and in Switzerland.

Table 12.8: **Average error in prediction at the decile level, by literacy scale and by country/language group (in percents)**

Prose		Document		Quantitative	
Country	Error	Country	Error	Country	Error
Sweden	1.4	Switzerland/French	2.2	United States	1.7
Germany	1.9	Germany	2.4	France	2.0
France	1.9	Netherlands	2.4	Switzerland/German	2.0
Poland	2.1	France	2.5	Sweden	2.2
United States	2.3	Sweden	2.6	Netherlands	2.5
Netherlands	2.4	United States	2.7	Germany	2.7
Switzerland/German	2.7	Poland	2.7	Poland	3.3
Switzerland/French	3.1	Switzerland/German	3.6	Switzerland/French	3.7
Canada/English	4.4	Canada/English	5.0	Canada/English	4.7
Canada/French	4.4	Canada/French	6.1	Canada/French	5.9

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

These results reflect something somewhat different from the multiple correlations: They assess how well the predicted and actual literacy levels agree *on a group basis* in comparing the deciles of risk. Stated another way, the goodness of fit analysis evaluates the evidence that the relationship between literacy and demographics is truly logistic in shape.

Tables 12.9 and 12.10 list the independent variables used in the logistic regressions and the code for each variable.

Table 12.9: **Independent variables used in the logistic regressions, by country/language group**

Country	Code	Independent variables used
Canada/English	1	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs2 lfs3 lfs4 lfs5 educ educ2 tv m_ed f_ed
Canada/French	2	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs3 lfs4 educ educ2 tv m_ed f_ed
France	10	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs3 lfs4 educ educ2 tv m_ed f_ed
Germany	5	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs2 lfs3 lfs4 lfs5 educ educ2 tv m_ed f_ed
Netherlands	8	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs2 lfs3 lfs4 lfs5 educ educ2 tv m_ed f_ed
Poland	9	sex age age2 inc inc2 lfs1 lfs3 lfs4 educ educ2
Sweden	11	sex age age2 imm inc inc2 lfs1 lfs2 lfs3 lfs4 lfs5 educ educ2 m_ed f_ed
Switzerland/German	3	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs2 lfs3 lfs4 lfs5 educ educ2 tv m_ed f_ed
Switzerland/French	4	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs3 lfs4 educ educ2 tv m_ed f_ed
United States	6	sex age age2 imm imyr imyr2 inc inc2 lfs1 lfs2 lfs3 lfs4 lfs5 educ educ2 tv m_ed f_ed

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Table 12.10: **Codes for independent variables used in the logistic regressions**

Code	Variable
sex	sex
age	age group
age2	(age group)^2
imm	immigrant (yes/no)
imyr	immigration year
inc	income quintile
inc2	(income quintile)^2
lfs1	employed
lfs2	retired
lfs3	unemployed
lfs4	student
lfs5	homemaker
educ	educational attainment
educ2	educational attainment
tv	television watching
m_ed	mother's educational attainment
f_ed	father's educational attainment

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The parameter coefficients for the various logistical regressions are presented in Appendix M. Standard errors are included in the tables because prediction, not statistical significance, is the main emphasis. Not all coefficients presented are significant.

For two reasons, it is difficult to compare the coefficient values for different countries. First, the variables selected for the regression equations were not always the same for the various countries. The use of quadratic terms means that interpreting the coefficients is awkward. For example, with age and age^2 present, a positive coefficient for the age variable alone does not indicate a positive correlation between age and the dependent variable. The coefficient of age^2 must also be taken into consideration. Second, the high level of colinearity in the independent variables means that the coefficients presented in the tables are unstable. This instability does not undermine the predicted values, however, which are the primary focus here.

12.5 Conclusions

It may be helpful to recall the interpretations of correlation and goodness of fit in the more familiar normal linear regression situation. The R^2 measures the extent of linear correlation. A moderately high R^2 or correlation does not necessarily imply a good fit of the data to a linear model, but a very high R^2 (over 90 percent) probably implies a linear model. In fact, it is entirely possible to have an R^2 of over 50 percent but a poor fit in the deciles of risk analysis. Similarly, it is also possible to have a good fit of the data to a linear model and have a very low R^2.

The results of the analyses described in this chapter show that the IALS countries would be ranked differently depending on the interpretation of predictability used, correlations or goodness of fit. Canada's data, for example, showed high individual correlations but a relatively poor fit. France's data, on the other hand, showed high individual correlations as well as a reasonably good fit using the decile information.

The analysis of R^2 values associated with the prediction of literacy levels suggests a fairly high level of prediction. The values ranged from about 40 to 60 percent for the various IALS countries.

The analysis based on goodness of fit concepts yielded very low average error terms. Many were in the order of 1 to 3 percent, with some as high as 6 percent. These error terms represent the average error observed when the population proportion at each literacy level for each decile of risk was predicted based on the observations.

Although the strength of the relationships vary from country to country, a common set of demographic characteristics can be used to predict the observed literacy skill profiles for the various countries with a reasonable degree of certainty. Thus, the IALS skill profiles are judged to be plausible estimates worthy of publication.

Chapter 13:

Nested-factor Models for the Swedish IALS Data

Bo Palaszewski

13.1 Introduction

Analyses of the IALS data performed by ETS assumed that the test items could be divided into three unidimensional scales: prose, document, and quantitative. As described earlier in this report, prose literacy refers to the ability to read and understand narrative texts, descriptions, and explanations. Document literacy refers to abilities related to understanding texts that consist of structured collections of information such as charts, graphs, lists, and tables. Quantitative literacy, or numeracy, refers to the ability to perform calculations using information given in texts or documents.

The assumption that these three literacy scales are unidimensional is fundamental to the IRT-based estimation technique used to scale the survey results using a common scale. This chapter scrutinizes the assumption of unidimensionality by using models for multivariate data. The analyses show that it is possible to identify a general literacy dimension as well as a more specific dimension measuring quantitative skill, hereafter referred to as numeracy. The more specific dimension could be due to context-specific factors within each set of items based on a common piece of text (Palaszewski 1997), or it may result from certain aspects of the study design (for example, speediness in the test).

The data for this special study were analyzed using a structural modeling approach which can accommodate the missing data introduced by the IALS design. A nested-factor model was chosen for the analyses. Gustafsson and Balke (1993) and Gustafsson (1995) define nested-factor models as models with different degrees of generality (e.g., a general factor) and more specific latent variables which are related to subsets of the manifest variables. Thus, there is a general factor which is related to all manifest variables, as well as other factor variables which are related to subsets of the manifest variables.

Research reported by Gustafsson (1994) has provided some limited support for using a hierarchical variance decomposition when reanalyzing the reading achievement data in the IEA study. The results from this study also suggested that it is reasonable to consider a model with a general reading achievement factor since the *text* and *document* factors were highly correlated.

The IALS analyses reported in this chapter assumed the existence of a general dimension, *literacy* (*Lit*), and the more specific dimensions *Doc* and *Num*. We also explored the existence of block effects. A special kind of block factor which accounts for test speediness, which we refer to here as an "end of test" factor, was of particular interest in the analyses.

13.2 Data

Because the number of test items in the IALS was very large, it was not possible to administer every test item to every subject. Therefore, a matrix sampling design was used. The items were divided into seven partially overlapping subsets, called blocks. Each block contained several sets of items, with each set based on a particular prose passage or document. The seven blocks were then compiled into seven test booklets, as indicated in Chapter 1, Table 1.3.

As shown in Table 13.1, each of the 2,919 persons who responded to the Swedish IALS completed one of the seven test booklets. In general, comparable numbers of individuals received each booklet. The number of subjects who received booklet 1 is artificially high because non-responding subjects were included in this group.

Table 13.1: **Number of subjects for each test booklet**

Booklet number	n
1	527
2	348
3	415
4	436
5	521
6	365
7	307

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

As a result of the matrix sampling design, each block of test items was completed by a subset of the examinees. The design ensured that the test was balanced with respect to the number of times each block occurred in the test booklets as a whole, as well as the location in which each block occurred; each block appeared exactly once in each position (first, second, and third).

What this design does not control for are the "carry-over" effects of the order in which blocks were administered to test subjects. In other words, a subject's exposure to a block of items may influence his or her motivation to answer the items in the following blocks.

13.3 Data Analysis

For this special study, the IALS data were analyzed by combining the seven subsamples (booklets), each with a somewhat different set of variables, and using techniques for analyzing structural equation models with missing data structures. These techniques were developed by Muthén, Kaplan, and Hollis (1987), Allison (1987), and others. In this case, the missing data are the result of the matrix sampling design described earlier. Gustafsson and Stahl's (1996) STREAMS system, combined with LISREL8 (Jöreskog and Sörbom 1993), makes it possible to implement the rather tedious data analyses required for this study.

Researchers analyzed structural equation models using 35 "passage scores" from the seven different blocks as manifest variables. Each passage score is the simple unweighted sum of scores (0 or 1) for the items corresponding to a particular passage or document. The use of passage scores was motivated by the fact that there is a local dependency between the items corresponding to each passage. In other words, the correctness of a response to an early item pertaining to a particular passage influences the probability of a correct answer for the following items related to that passage. Thus, using passage scores rather than item scores makes the analyses more robust. It also minimizes the number of variables, an important consideration given the computer-intensive nature of data analyses based on matrix sampling designs.[1]

The data analyses were based on the assumption that a common literacy dimension exists for all passages. The use of the *Lit* dimension as this general dimension was partly motivated by results reported in Palaszewski (1997). The *Num* dimension was related to all passages that included at least one item classified as measuring quantitative abilities, and the *Doc* dimension was related to passages that included at least one item measuring the ability to interpret documents.

As shown in Table 13.2, this preliminary analysis shows that a model with *Lit* as a general factor and *Num* and *Doc* as specific factors fits the data rather well.

Table 13.2: **Test values for all subjects**

Model	x^2	d.f.
Lit	2,609.6	560
Lit+Num	2,288.0	540
Lit+Num+Doc	2,027.0	520

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Next, the subjects were divided into two groups, male and female, and multigroup analyses were performed. The results are presented in Table 13.3. Models were first estimated with parameters restricted to be equal for males and females. The factors for these models appear in parentheses. Models were then estimated with parameters free of such restrictions. The factors for these models appear without parentheses.

---

[1] In the passage score analysis reported here, not-reached items were scored as incorrect. This approach differs somewhat from the methods used to generate proficiency estimates in the IALS, reported in Chapter 11.

The best fitting model has common parameter estimates between males and females in the Lit dimension but unconstrained parameter estimates between the two groups in the Num dimension.

Table 13.3: **Test values for multigroup analysis by sex**

Model	x^2	d.f.
(Lit)	3,828.1	1,225
Lit	3,765.6	1,191
(Lit)+(Num)	3,506.8	1,204
(Lit)+Num	3,455.2	1,186
(Lit)+Num+(Doc)	3,452.3	1,185
(Lit)+Num+Doc	Non-convergence	

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

In this analysis, the effect of the Doc dimension is no longer significant. This could be interpreted as evidence of a sex-specific mean difference in this dimension. The previously significant factor estimate for this dimension could then be explained as an aggregation effect, and not as a real factor effect. Another plausible explanation of the sex-specific mean difference is given by Rosén (1995). She found that in the IEA data, a difference between males and females in the Doc factor was mainly due to context-specific contributions. This could also be true in the IALS study.

To explore possible block effects, each test booklet was analyzed separately.[2] For every book, a model was fitted including the Lit factor as a general dimension and a block factor accounting for a block of "passages." The parameter estimates within a block were restricted to be equal. The block estimates are indicated as (Blockn), where Blockn represents all passages within the n:th block.

The blocks were introduced one at a time into the model, and this procedure was repeated for each block within a booklet. This process yielded significant block effects which could be interpreted as the result of speediness. Speediness refers to the extent to which items late in the test were completed adequately. For the most part, the blocks in which significant block effects were found appeared in the final position in the test booklets. Six of the significant block effects were found among the blocks in the final position. Three of the remaining significant block effects were for blocks in position 1, and another three were for blocks in position 2. Thus, there seems to be some limited support for the existence of a speediness factor.

In the full model, in which data from all books are combined in a common missing data model, existing block effects for block factors (and speediness factors) are averaged out to some extent, since data with block effects are weighted together with data from blocks without block effects. Still, the full model, with block effects included, yielded a significantly better fit.

---

[2] It should be noted that the seven IALS blocks were not designed to be psychometrically identical to one another. The blocks differ in difficulty, in length, and in the distribution of items from the three literacy scales. Thus, differences among the blocks would not have a negative impact on the proficiency estimates.

The results in Table 13.4 were generated by restricting estimates for the Lit dimension to be equal for males and females and by allowing for block effects.

Table 13.4: **Tests for block effects**

Model	x^2	d.f.
(Lit)	3,828.1	1,225
+(Block1)+	3,807.5	1,224
+(Block2)+	3,806.3	1,223
+(Block3)+	3,725.9	1,222
+(Block4)+	3,708.9	1,221
+(Block5)+	3,700.6	1,220
+(Block6)+	3,700.5	1,219
+(Block7)	3,666.9	1,218

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

Every addition of a block factor may be tested by taking its x^2 value and subtracting the x^2 value from the previous model. Significant block factors were found for all blocks except Blocks 2 and 6. Tests were also conducted to determine whether there were any significant differences in block effects between males and females. The results in Table 13.5 show that significant sex differences were detected only for Block 3.

Table 13.5: **Tests for block effects by sex**

Model	x^2	d.f.
+Block1+	3,666.9	1,217
+Block3+	3,638.7	1,216
+Block4+	3,638.2	1,215
+Block5+	3,637.8	1,214
+Block7+	3,637.7	1,213

Source: Statistics Canada and Educational Testing Service, International Adult Literacy Survey, 1994.

The best fitting model for the Swedish IALS data was a model with a general Lit factor constrained across sex groups plus block factors for Blocks 1, 3, 4, 5, and 7. For the block factors, separate parameters were fitted for each sex only for Block 3. A Num factor was included with unrestricted parameter estimates between sex groups. According to our previously used notation, the model could be written as:

(Lit)+ (Block1)+Block3+ (Block4)+ (Block5)+ (Block7) +Num

The x^2 for this model was 3,200.1 with 1,179 degrees of freedom.

Parameter estimates for this model are included in the appendices to this report. For the Num dimension, only the parameter estimates for males were significant. Most of these estimates (13 out of 20) were positive; only 4 had a negative sign of the factor coefficient. This pattern of differentiation in the Num dimension—that is, males varied in numeracy skills, but females did not—might be explained by differences in occupational content or training. For example, exposure to numerical job content may vary more among males than among females.

13.4 Conclusions

This chapter provides preliminary evidence concerning factors that underlie the literacy proficiencies of Swedish adults tested in the IALS. While far from complete, the analyses strongly suggest a reasonable regularity to the observed proficiency distributions.

The final nested-factor model fitted on the passage scores obtained a reasonably good fit. The general dimension measured by the Lit factor explained most of the variability in results, although a numerical dimension also appeared to exist. The general dimension did not differ by sex, but mean differences were found between males and females in the numerical dimension. These differences could be explained by the existence of gender-specific content in the assessment instrument. The numerical dimension measured as a factor could only be established for the male group. This pattern of differentiation in numeracy among males could be a result of differences in job content between males and females.

References

Allison, P.D. 1987. Estimation of linear models with incomplete data. In C. Clogg (Ed.), *Sociological methodology 1987*. San Francisco: Jossey Bass.

Balke, G. 1995. Decomposition of reading comprehension. Analysis of IEA reading literacy tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Bassett, E.D., Goodman, D.G., and Fosegan, J.S. 1981. *Business records controls* (5th ed.). Cincinnati, OH: South-Western Publishing.

Beaton, A. E., and Johnson, E. G. 1990. The average response method of scaling. *Journal of Educational Statistics, 15*, 9-38.

Birnbaum, A. 1968. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bishop, J.H. 1995. Improving education: How large are the benefits? How can it be done efficiently? (Working paper #95) Ithaca, NY: Center on the Educational Quality of the Workforce, Cornell University.

Bock, R.D. and Aitkin, M. 1981. Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika, 46*, 443-449.

Braun, H.I. and Holland, P.W. 1982. Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P.W. Holland and D.B. Rubin (Eds.), *Test equating*. New York, NY: Academic Press.

Burch, J.G., and Grudnitski, G. 1986. *Information systems: Theory and practice* (4th ed.). New York: Wiley.

Clark, H.H. and Haviland, S.E. 1977. Comprehension and the given-new contract. In R.O. Freedle (ed.), *Discourse processes: Advances in theory and research* (vol. 1), pp. 1-40. Norwood, NJ: Ablex.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society (Series B), 39*, 1-38.

Divgi, D.R. 1985. A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement, 9*, 413-415.

Fisher, D. 1981. Functional literacy tests: A model of question-answering and an analysis of errors. *Reading Research Quarterly*, 16, 18-8.

Fry, E. 1977. Fry's readability graph. Clarifications, validity, and extension to level 17. *Journal of Reading, 21*, 242-52.

Gustafsson, J.E. 1995. Alternative hierarchical models of reading achievement. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Gustafsson, J.E. and Balke, G. 1993. General and specific abilities as predictors of school achievement. *Multivariate Behavioural Research, 28*, 407-434.

Gustafsson, J.E. and Stahl, P.A. 1996. *STREAMS user's guide*, version 1.6. Molndal, Sweden: Multivariate Ware.

Guthrie, J.T., Seifert, M., and Kirsch, I.S. 1986. Effects of education, occupation, and setting on reading practices. *American Educational Research Journal, 79*, 220-27.

Haebara, T. 1980. Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research, 22*, 144-149.

Holland, P. 1985, October. On the study of differential item performance without IRT. *Proceedings of the military testing association.*

Holland, P. and Thayer, D. 1988. Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H.I. Braun (Eds.), *Test validity*. Hillsdale, NJ: Erlbaum.

Johnson, E.G., Mislevy, R.J. and Zwick, R. 1990. Estimation of the standard errors of the adjusted 1986 NAEP results. In A. E. Beaton and R. Zwick, *Disentangling the NAEP 1985-86 reading anomaly*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Johnson, E.G., and Rust, K.F. 1992. Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics.*

Jöreskog, K.G. and Sörbom, D. 1993. LISREL8 reference guide. Chicago: Scientific Software International.

Kirsch, I.S., and Jungeblut, A. 1986. *Literacy: Profiles of America's young adults* (NAEP report no. 16-PL-01). Princeton, NJ: Educational Testing Service.

Kirsch, I.S., and Jungeblut, A. 1992. *Profiling the literacy proficiencies of JTPA and ES/UI populations: Final report to the Department of Labor*. Princeton, NJ: Educational Testing Service.

Kirsch, I.S., Jungeblut, A., and Campbell, A. 1992. *Beyond the school doors: The literacy needs of job seekers served by the U.S. Department of Labor.* Princeton, NJ: Educational Testing Service.

Kirsch, I.S., Jungeblut, A., Jenkins, L., and Kolstad, A. 1993. *Adult literacy in America: A first look at results from the National Adult Literacy Survey*. U.S. Department of Education: National Center for Education Statistics.

Kirsch, I.S. and Mosenthal, P.B. 1989. Building documents by combining simple lists. *Journal of Reading, 33*, 132-5.

Kirsch, I.S. and Mosenthal, P.B. 1990a. Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*, 5-30.

Kirsch, I.S. and Mosenthal, P.B. 1990b. Nested lists. *Journal of Reading, 33*, 294-7.

Kirsch, I.S. and Mosenthal, P.B. 1992. How to navigate a document using locate know/need-to-know strategies. *Journal of Reading, 36*, 140-3.

Levine, R.S. 1955. *Equating the score scales of alternate forms administered to samples of different ability* (ETS Research Bulletin 55-23). Princeton, NJ: Educational Testing Service.

Linn, R.L., Levine, M.V., Hastings, C.N., and Wardrop, J.L. 1981. An investigation of item bias in a test of reading comprehension. *Applied Psychological Measurement, 5*, 159-173.

Little, R.J.A. and Rubin, D.B. 1983. On jointly estimating parameters and missing data. *American Statistician, 37*, 218-220.

Lord, F.M. 1980. *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Marco, G.L. 1977. Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement, 14*, 139-160.

Messick, S. 1989. Validity. In Robert L. Linn (ed.), *Educational Measurement* (3rd ed.). New York: Macmillan.

Mislevy, R.J. 1984. Estimating latent distributions. *Psychometrika, 49*, 359-381.

Mislevy, R.J. 1985. Estimation of latent group effects. *Journal of the American Statistical Association, 80*, 993-97.

Mislevy, R.J. 1990. Scaling procedures. In E.G. Johnson and R. Zwick, *Focusing the new design: the NAEP 1988 technical report* (No. 19-TR-20). Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Mislevy, R.J. 1991. Randomization-based inferences about latent variables from complex samples. *Psychometrika, 56*, 177-196.

Mislevy, R.J., Beaton, A., Kaplan, B.A., and Sheehan, K. 1992. Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29(2)*. 133-161.

Mislevy, R.J. and Bock, R.D. 1982. *BILOG: Item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software.

Mislevy, R.J., Johnson, E. G., and Muraki, E. 1992. Scaling procedures in NAEP. *Journal of Educational Statistics, 17-2*, 131-54.

Mislevy, R.J. and Sheehan, K.M. 1987. Marginal estimation procedures. In A. E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (no. 15-TR-20), pp. 293-360. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Mislevy, R.J. and Sheehan, K.M. 1989. The role of collateral information about examinees in item parameter estimation. *Psychometrika, 54*, 661-679.

Mohadjer, L., Berlin, M., Reiger, S., Waksberg, J., Rock, D., Yamamoto, K., Kirsch, I.S., and Kolstad, A. 1994. The role of incentives in literacy survey research. Paper presented at the PECD/NCAL Roundtable on Adult Basic Skills: Innovations in Measurement and Policy Analysis, Philadelphia, PA.

Mosenthal, P.B., and Kirsch, I.S. 1989a. Lists: The building blocks of documents. *Journal of Reading, 33*, 58-60.

Mosenthal, P.B., and Kirsch, I.S. 1989b. Intersecting lists. *Journal of Reading, 33*, 210-3.

Mosenthal, P.B., and Kirsch, I.S. 1991. Toward an explanatory model of document literacy. *Discourse Processes, 14*, 147-80.

Muthén, B., Kaplan, D., and Hollis, M. 1987. On structural equation modelling with data that are not missing completely at random. *Psychometrics, 52*, 431-462.

OECD. 1992. *Adult illiteracy and economic performance*. Paris: OECD.

OECD. 1996. Lifelong learning for all. Meeting of the OECD education ministers, 16-17 January 1996.

Palaszewski, B. 1997. Multidimensional models for matrix-sampled reading tasks. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Petersen, N.S., Kolen, M.J., and Hoover, H.D. 1989. Scaling, norming, and equating. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.). New York, NY: Macmillan.

Rosén, M. 1995. *Gender differences in performance on documents as a function of cultural factors*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Rubin, D.B. 1987. *Multiple imputation for nonresponse in surveys*. New York: John Wiley and Sons.

Sheehan, K.M., and Mislevy, R.J. 1988. *Some consequences of the uncertainty in IRT linking procedures* (RR-88-38-ONR). Princeton, NJ: Educational Testing Service.

Shillington, R. 1996. *Estimating literacy scores using least-squares regression and logistic regression analysis*. Unpublished manuscript.

Statistics Canada. 1990. *An International Assessment of Adult Literacy: A proposal*. Unpublished manuscript.

Statistics Canada. 1991. *Adult literacy in Canada: Results of a national study*. Ottawa, ON: Statistics Canada.

Statistics Canada and OECD. 1996. *Literacy, economy and society: First results from the International Adult Literacy Survey*.

Statistics Canada, Human Resources Development Canada, and OECD. (in progress). *Literacy Skills for the Knowledge Society: Further Results of the International Adult Literacy Survey*.

Stocking, M.L. and Lord, F.M. 1983. Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201-210.

Thomas, N. 1993. Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics*, 2, 309-22.

Tsutakawa, R.K., and Johnson, J. 1990. The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika, 55*, 371-390.

United States Department of Education. (in progress). *Technical report on the National Adult Literacy Survey*. National Center for Education Statistics.

Vale, C.D. 1986. Linking item parameters onto a common scale. *Applied Psychological Measurement, 10*, 333-344.

Way, D.W. and Tang, K.L. 1991. *A comparison of four logistic model equating methods*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Wickert, R. 1989. *No single measure: A survey of Australian adult literacy*. The Commonwealth Department of Employment, Education, and Training, Government of Australia.

Wingersky, M, Kaplan, B.A., and Beaton, A.E. 1987. Joint estimation procedures. In A.E. Beaton, *Implementing the new design: The NAEP 1983-84 technical report* (No. 15-TR-20), pp. 285-92. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Yamamoto, K. and Muraki, E. 1991. *Non-linear transformation of IRT scale to account for the effect of non-normal ability distribution on item parameter estimation*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Appendix A:

Review of Methodology

Graham Kalton
Lars Lyberg
Jean-Michel Rempp

1. Introduction

The International Adult Literacy Survey (IALS) is the first large-scale comparative assessment of adult literacy skills ever undertaken. The survey has been conducted in eight countries in the first round, a second round of the survey in another four countries is now being developed, and a further round is under negotiation with the European Union and OECD. In one of the countries in the first round concerns have been raised about the international comparability of the survey data. To address these concerns, Statistics Canada, as the coordinator of the first round of the survey, decided that the IALS methodology should be subjected to an external evaluation. Statistics Canada therefore invited the three of us who work as survey methodologists in three of the participating countries but who have had no previous involvement in any phase of the IALS to undertake an independent review of the IALS methodology. We were informed that the IALS would follow any recommendations made by us even if that would entail the withdrawal of all the study results.

Our work started in August 1995 and this is our report. We met with IALS coordinators in Paris, September 17-19, and in London, October 14-15. In London we also presented our preliminary findings to the national study directors. During the period September-November we have collaborated with each other and with the IALS coordinators at Statistics Canada. We have also had meetings and correspondence with the national study directors. Statistics Canada and the national survey organizations have been extremely cooperative, providing all the information we have requested.

Our approach has been to collect information on the survey procedures employed in the eight countries participating in the first round of the IALS. To this end, we developed a questionnaire that covered the methodological issues that we considered important. In most cases the questionnaire for a particular country was completed by one of us on a visit to the national survey organization. The information collected in this process is reviewed in later sections of the report.

It should be noted that time and resources did not permit us to carry out any analyses of the survey data. Our report can therefore only point out where survey procedural errors occurred. We have not been able to investigate what the likely consequences of such errors might be. As noted in the final section, one of our recommendations is that further work should be conducted to provide empirical evidence on the possible consequences of the various sources of survey error for the international comparability of the IALS data.

2. Overview of Findings

Multi-national surveys are important instruments for advancing the field of social and other sciences. The design, execution, and analysis of such surveys present severe challenges, however. Ideally, in multi-national surveys sampling strategies, data collection procedures, and coding and analysis procedures should be kept uniform so that comparability can be achieved. Many factors work against this ideal scenario. Clearly in the field of adult literacy, the development of measurement instruments in different languages that maintain conceptual and/or literal meaning across nations is a difficult task. Even when such instruments are developed, there remain problems of securing standardization in other aspects of survey methodology. In particular differing cultures of survey practice across nations and across survey organizations present difficulties in securing standardized procedures. Also, when the survey funds are raised individually by each country, differing funding levels and sometimes insufficient funding are detrimental to the achievement of high-quality standardized procedures.

Based on the information we collected, we believe that the considerable efforts that were made to develop standardized survey instruments, such as the background questionnaires and exercise booklets, for the different nations and languages were successful, and that the data obtained from them should be broadly comparable. Thus, we do not believe that the measurement procedures should be seriously questioned. With regard to other aspects of the survey methodology, however, the standardization of procedures has not been achieved to the extent desired. The IALS coordinators provided guidance on survey procedures to the participating countries, but did not enforce adherence to specific procedures. As a result, participating countries had a fair degree of latitude in the design and conduct of their national surveys. Nonresponse has proved to be a particular weakness, with generally very high nonresponse rates and variation in nonresponse adjustment procedures across countries. For some countries the sample design is problematic resulting in some unknown biases. The data collection and its supervision differ between participating countries, and some clear weaknesses are evident for some countries.

The variation in survey execution across countries is so large that we recommend that all comparative analyses across countries should be interpreted with due caution. In particular, we recommend against the publication of comparisons of overall national literacy levels. We consider any ranking of countries based on such comparisons to be of dubious value given the methodological weaknesses referred to above.

Despite the methodological weaknesses, we recommend that the survey results be published. The instruments developed for measuring adult literacy constitute an important advance, and the results obtained from the instruments in the first round of the IALS are a valuable contribution to this field. We recommend that the survey report focusses on analyses of the correlates of literacy (e.g., education, occupation, and age) and the comparison of these correlates across countries. Although these analyses may also be distorted by methodological problems, we believe that they are likely to be less affected by these problems than are the overall literacy levels.

In making our recommendations, we recognize that we may be setting higher standards than have been employed in some previous international surveys. However, as trained survey methodologists, we are seriously concerned about the possible effects that the deviations from sound survey practice that we have observed may have on the survey results. We think that the results of the first round of the IALS should be accompanied by warnings of the possible biases that may exist.

Some aspects of the first round of the IALS have worked well. In particular, the instruments and psychometric analyses appear to have worked well and the coverage of the adult population in each country has been good. However, the experience obtained in the first round of the survey has identified areas of survey practice where improvements are needed. In the spirit of continuous quality improvement, we believe that advantage should be taken of that experience to develop improved methods for use in future rounds of the survey. Some recommendations for improvements are presented in Section 11.

3. Potential Error Sources

The first round of the IALS has been conducted in Canada, France, Germany, the Netherlands, Poland, Sweden, Switzerland, and the U.S.A. Moreover, the survey has been conducted in two languages in Canada (English and French) and Switzerland (French and German). One of the aims of the IALS is to institute standardized survey designs in all the participating countries so that the resultant data can be analyzed as a whole and so that inter-country comparisons of adult literacy can be made. The following sections of this report review key aspects of the methodology used in the eight participating countries in order to assess the comparability of the survey data across countries and to identify any potential biases that might be present. It is important to stress the word potential. We have not had the time to attempt to assess the magnitude of any biases. We can only judge the designs and the execution of the surveys based on the various sources of information available to us, including the responses to the questionnaire that we administered to the national survey organizations involved in executing the IALS.

A critical issue for all international surveys is that the questionnaires and other measuring instruments used in the participating countries produce comparable data. In part, this relates to the need to translate instruments into equivalent forms in the different languages but, more than that, the instruments need to conform to the different cultural settings found in the various countries. The task of producing comparable measures of adult literacy across countries is a particularly challenging one. The efforts to produce comparable measures are evaluated in Section 4.

Comparability in the measuring instruments is a necessary but not sufficient condition for comparability of data in multi-national surveys. To fully evaluate data comparability, consideration must also be given to other survey design aspects. In particular, the following features of the surveys need to be evaluated:

Sample design. Strict probability sampling methods were applied in some but not all countries participating in the first round of the IALS. When deviations from probability sampling occur, there is the risk of biases in the survey estimates. Section 5 reviews the sample designs employed in the various countries.

Coverage. An important consideration is the population covered by the sample designs for the various countries, for example, the age range of the population surveyed and whether foreigners are included. Section 6 reviews the coverage in the various countries.

Nonresponse. The IALS involves a substantial degree of respondent burden and its contents are threatening to some respondents. These factors make the task of gaining the cooperation of respondents to participate in the survey a challenging one. In the first round of the IALS the nonresponse rates are so substantial that resulting biases in survey estimates are probable. Section 7 reviews the nonresponse rates obtained in the participating countries and the efforts made to secure response.

Data collection procedures. The quality of the IALS data depends on the data collection procedures used. Important factors include the experience and training of interviewers, their supervision, quality checks performed on their work, the way the survey is introduced to the sampled persons, and the procedures used for the interviews. Section 8 reviews these procedures.

Data processing and data entry. Coding in this survey includes the coding of industry, occupation and education in the background questionnaire and the coding of the responses to the test items in accordance with the international standards established between all the IALS countries. The coding needs to be performed in a consistent way across countries. The survey data have to be entered into the computer files for analysis, and errors can and do occur at this stage as well as at the coding stage. Section 9 reviews the steps taken to ensure coding consistency and to control the quality of data entry.

Weighting. The data from respondents need to be weighted in the analysis to compensate for the use of unequal selection probabilities. Weighting adjustments may also be used to attempt to compensate for nonresponse and noncoverage. Section 10 comments on this issue.

Section 11 presents our five major recommendations based on the results of our study. These recommendations are of two types. Three relate to the first round of the IALS; they concern the publication of the survey results and the need for further methodological evaluation. The other two make recommendations for the future rounds of the IALS.

It should be noted that our study has been concerned only with the issue of data comparability across the participating countries. There is another aspect of comparability that is beyond the scope of our report. This aspect concerns the statistical analyses of the data and the interpretation of the findings. Even if the data are reasonably comparable, the analyses and their interpretations need to be sensitive to the differences between countries. For example, an insightful comparison of reading scores between countries needs to take into account the size of the adult immigrant population in each country. To the extent that the native languages of immigrants are not those of the country in which they reside, this can be expected to lead to lower reading scores for immigrants, and hence for the country. As this example shows, differences between the average reading scores of different countries cannot be simply explained by differences in educational practices. Rather, in-depth inter-country comparisons require sophisticated analyses that take many factors into account. This issue is, however, not treated here.

4. Questionnaires and Measurements

Considerable efforts were devoted in the IALS to securing measurement conformity across countries. Each country had to translate the task booklets, task administration guides, and the scoring rubrics. The desire to make the items as culturally neutral as possible within each country made some adaptation necessary in addition to the translation. A list of allowable alterations to each item was provided and each country's semifinal documents were reviewed by Statistics Canada and a panel of linguists. The review resulted in a number of recommendations for change for the individual countries. Statistics Canada also reviewed the background questionnaires and offered suggestions to each country.

In order to ensure that the final test design included only those items having a high probability of functioning, each country was required to conduct a pilot survey. The choice of well performing items for inclusion was approved by participating countries. Overall, it appears that a very high degree of conformity in the test items has been achieved across countries and languages. Nevertheless, we did note a few differences between countries, such as differences in layout (e.g., a key word underlined in one country but not in another) and in content (e.g., the use of the same word in the question and the document to which it referred in one country but the use of different words in another country). We believe that these differences are unlikely to have a noticeable effect on the general results.

Each country agreed to rescore a minimum of 20% of all assessments to detect and correct scorer variability and scoring drift. Countries also exchanged booklets when language so permitted, in order to control scoring consistency between countries. The results show a relatively small number of between country scoring problems and high levels of confidence within the country scoring process.

We were not able to conduct a detailed review of the psychometric methods used for the analysis of the test results, but from our cursory examination the methods appear to be sound. The majority of the items appear to scale consistently across countries. As a rough check, we compared the country rankings based on the psychometric scores with those based on crude proportions of items answered correctly for each of the scales. These comparisons showed that the rankings were in broad agreement for all three scales.

Our overall assessment is that the development of questionnaires and the quality of the psychometric analysis are very stable components of the IALS study and they do not give rise to any major concern.

5. Sample Design

The design specifications stated that each participating country should use a high quality representative sample of individuals. Although both "high quality" and "representative" are features that could be interpreted in different ways, we believe that probability samples should be used. Unfortunately, some of the sampling designs used in the first round of the IALS do not satisfy the conditions for probability sampling.

The essential conditions for probability sampling are that the sample selection probabilities are known and that every member of the target population has a non-zero probability of being included in the sample. If these conditions are satisfied, unbiased (or at least approximately unbiased) estimates of the population parameters can be computed from the sample data. If probability sampling is used, there is no need to require the use of the same sample design in each country. Rather, the designs may be constructed in the most efficient manner for each country, making use of the data sources available in the country. If probability sampling is not used, the survey estimates are subject to biases of unknown magnitude, and these biases make attempts to compare survey results across countries hazardous.

If probability sampling is used, there remains the issue of whether the sample design is an efficient one for the purposes of the survey. An inefficient probability sample design will produce less precise estimates than necessary, and this fact will be reflected in the standard errors of the estimates. An inefficient probability sample will not result in biased estimates. Since our primary focus is on the comparability of estimates across countries, we are more concerned here with potential sampling bias than with the precision of the estimates.

With this background, we now list the basic features of the sample designs in the participating countries followed by some comments:

Canada. Canada used two samples. The main IALS sample was a subsample of the May 1994 Canadian Labour Force Survey (LFS) file using probability sampling at all stages. The subsample was a stratified sample of 6,427 LFS respondents, with an oversample of certain target groups of policy interest. The second sample was a three-stage probability sample of Francophones from the province of Ontario selected from the 1991 Census.

France. France's sample was selected using a three-stage sample design of households. The last stage involved a random route strategy. The rules for the random route selections were relatively complex and were not always explicitly written in the interviewers' instructions. In communities of more than 10,000 inhabitants, the sample of households was not selected by the interviewers. The initial sample size was about 6,000 households (including approximately 20% of addresses in which there was no eligible person). In each selected household, the birthday method was used for selecting the person to be interviewed. The final sample size was 2,996.

Germany. Germany employed a master sample of sampling points, with the selection of addresses being made using the random route method. At each of the 525 sampling points, a single random route of 23 addresses was followed, and in each of them exactly eight addresses were selected. In each household the interviewed person was selected by the Kish method. The sample comprised 4,033 addresses, of which 997 were out-of-scope, and 2,100 interviews were conducted.

The Netherlands. The Dutch approach was two-stage systematic sampling. In the first stage postal codes were selected and in the second one address was selected per selected postal code. The birthday method was used for selecting the person to be interviewed in each sampled household. It took a sample size of 8,191 addresses to obtain 3,000 respondents.

Poland. Poland used a stratified, multi-stage design employing probability sampling at the various stages. The sample was selected from the Polish National Register of Citizens, a register that covers all persons living permanently (for longer than three months) in Poland. Data were collected for 3,000 persons.

Sweden. Sweden used a stratified master sample that was self-weighting. The sample was drawn from a national register of individuals. An initial sample size of 5,275 persons resulted in 3,038 interviews.

Switzerland. The target population was divided into two strata, corresponding to German- and French-speaking regions. Household telephone numbers were selected, and in each household the first eligible member in the alphabetical order was selected. A complementary sample was selected in the canton of Geneva using the same methods as the principal sample. The initial sample size comprised more than 7,100 telephone numbers (including out-of-scope households). Interviews were conducted with 2,843 persons, of whom 1,399 were German speaking and 1,444 French speaking.

U.S.A. The U.S.A. selected its sample from housing units undergoing their final Current Population Survey interviews in March-June, 1994. A probability sample of 4,901 persons was selected using a disproportionate stratified design, with strata formed by race/ethnicity and education. The disproportionate allocation was designed to provide an efficient psychometric linkage of the IALS survey to the earlier National Adult Literacy Survey (NALS).

The sample designs of France and Germany are problematic because of the random route component. Since selection probabilities are not known with random route sampling, the French and German samples are not probability samples. In Germany it is clear that the probabilities of inclusion are not equal between households, because the probabilities of selecting the possible starting points for the random routes were very unequal. A major issue with random route designs is the control of the sample selection in the field, with a serious concern that the interviewers exert an influence on the household selections. Control of the selection process requires detailed instructions that are faithfully followed by the interviewers. The instructions for the interviewers in France were very complex. The sample designs employed in France and Germany might be associated with a bias that is impossible for us to assess.

In Switzerland the sample of households was selected from the telephone directory without taking nontelephone households and special nonlisted groups of persons into account. The fact that some households have more than one telephone number is not considered except for the statement that that situation is not very common. The selection of the first eligible person in the alphabetical order within a household is also problematic, since this makes the sample a nonprobability one.

We might conclude that all countries except France, Germany, and Switzerland have sampling designs with known non-zero inclusion probabilities. The French, German, and Swiss designs do not have this feature. However, that does not necessarily mean that their sample estimates are seriously biased; that we cannot assess. In future rounds of the IALS we recommend that strict probability sampling be required.

6. Coverage

The design specifications for the IALS stated that in each country the study should cover the civilian, noninstitutional population aged 16-64. It is usual practice to exclude the institutional population from national surveys because of the difficulties in conducting interviews in institutional settings. Similarly, it is not uncommon to exclude certain other parts of a country's population that pose difficult survey problems (e.g., persons living in sparsely populated areas). The following is a list of the exclusions and of the intended coverage rates of the total population where available:

Canada. Exclusions include residents of specific territories, members of the armed forces and residents of institutions. The intended coverage rate is approximately 98%.

France. Exclusions include Corsica (0.5% of the population), overseas regions (2.5%) and residents of institutions (2%) - collective workers homes, students' dormitories, old people's homes, prisons, and barracks. Students who are living in dormitories and conscripts called up for national service are counted at their home address (usually their parents' address), if they have one.

Germany. Many surveys in Germany exclude foreigners, but foreigners were included in the IALS. Interviewers were given special instructions to include foreigners.

The Netherlands. Persons living in institutional dwellings are excluded. The intended coverage rate is approximately 99%.

Poland. Persons who have resided in Poland for less than three months are not included on the sampling frame. The number of such persons is small.

Sweden. Exclusions include non-Swedish speaking persons, persons living abroad during the whole survey period, and persons living in institutions.

Switzerland. The Italian and Rheto-romanic speaking areas of Switzerland were excluded. They contain about 4.8% of the Swiss population. The population living in institutions was also excluded (3.5% of the population). The use of the telephone directory as the sole frame presents another type of coverage problem: it excludes nontelephone households and households with unlisted telephone numbers (approximately 2%).

U.S.A. Exclusions include members of the armed forces, residents of institutions, homeless persons and persons living in housing units constructed after the selection of the CPS sample from which the IALS sample was drawn. The last category probably comprises about 2.5 to 3.5 percent of the population.

The intended coverage of the surveys generally conforms well to the design specifications. It should be noted, however, that the above discussion relates to the *intended* coverage of the surveys, where the exclusions are made by design. The actual coverage is generally lower than the intended coverage because of deficiencies in sampling frames and sampling frame construction (e.g., failures to list some households and some adults within listed households). An indication of actual coverage may be obtained by comparing population sizes estimated from the survey with external benchmark figures, both for the total population and for population subgroups. In the United States, for example, such comparisons suggest that the overall coverage rate for the Current Population Survey (the survey from which the IALS sample was selected) is about 93%, but that it is much lower for certain population subgroups (particularly young Black male adults).

We did not have the time and resources to carry out an investigation to compare population estimates from the surveys with benchmark estimates for the various countries. However, we believe that a thorough investigation of this type should be conducted, and we recommend that this be done as part of a further methodological evaluation of the IALS surveys.

7. Nonresponse Rates

The design specifications stated that the overall nonresponse rate should be less than 20%. In fact, the requirement was even more stringent since, in order to minimize the risk of nonresponse bias, this requirement also needed to be satisfied for all important strata. To be counted as respondents, persons had to complete a substantial portion of the background questionnaire.

No country met the nonresponse rate requirements, and most fell well short of them. The definitions of nonresponse rates may differ between countries, and it has not been possible for us to fully identify these differences. We cannot therefore confidently report nonresponse rates that are exactly comparable across countries. However, we believe that the nonresponse rates given below are reasonably comparable.

Canada. The IALS sample in Canada was a subsample of the Labour Force Survey (LFS). Allowing for both nonresponse to the LFS and to the IALS, the overall nonresponse rate is about 32.4%.

France. The exact number of out-of-scope addresses is not known, but it seems that the nonresponse rate is approximately 40% of the eligible addresses. The French study design involved a relatively complex and biased system for replacements, and the breakdown of nonresponses into categories was not complete.

Germany. The nonresponse rate is 31%.

The Netherlands. The nonresponse rate is around 55% which might be an underestimate. The fact that no interpreters were used to obtain background information from the non-Dutch speaking persons could add to that figure.

Poland. The initial nonresponse rate is approximately 25%. The handling of replacements makes the calculation of final nonresponse rate unclear.

Sweden. The nonresponse rate is 40.5%.

Switzerland. The nonresponse rate is approximately 45%.

U.S.A. The IALS sample in the U.S.A. was a subsample of the Current Population Survey (CPS). Allowing for a 5% nonresponse rate for the CPS, the overall nonresponse rate is approximately 41%.

These high nonresponse rates are disturbing. It appears that a number of countries have focussed on achieving the desired number of respondents rather than on securing high response rates. Unfortunately, studies with nonresponse rates like the ones listed above can be challenged by anyone who cares to do so on the grounds that their results may be subject to serious biases. We have information on sources of nonresponse but we are not in a position to assess potential biases.

The Swedish IALS sponsors were concerned about the large nonresponse rate in the Swedish survey and conducted a small evaluation study. Preliminary results indicate that the original nonrespondents who participated in the evaluation study were persons with very high and very low levels of education, with few persons in the middle. It is interesting to note that half of the original nonrespondents agreed to participate in the evaluation study without extensive efforts on the part of the survey managers.

In some countries it is possible to obtain useful information about the IALS nonrespondents from the sampling frame from which they were selected or from the original survey from which the IALS sample was selected. In Canada and the U.S.A., for instance, a great deal of information about the IALS nonrespondents can be obtained from the information collected in the Labour Force Survey and Current Population Survey, respectively. We recommend that, where possible, analyses be conducted to use such information to determine the characteristics of IALS nonrespondents. The results of these analyses may also be used in the development of revised weighting adjustments that attempt to compensate for the nonrespondents.

We believe that much more should have been done to meet the response rate requirements. In many cases more effort should have been made to contact the not-at-homes, to convert refusals, and to trace and contact movers (where applicable). A high-quality fieldwork organization with well-trained, experienced interviewers and a strong supervisory organization is necessary to secure high response rates in a complex survey like the IALS. We believe that it is essential that future studies enforce some standardized current best methods to deal with nonresponse. Elements of such a practice might include rules for tracing and refusal conversion, systematic collection of data on nonrespondents, evaluation studies and appropriate weighting procedures.

8. Data Collection

According to the design specifications countries should, to the extent possible, adhere to the training and collection procedures set out in the models provided by Statistics Canada. Major deviations from the standard approach needed to be approved by Statistics Canada in advance.

There is a fair degree of variation in the data collection procedures used in different countries. In all but one country, interviewer training has included special training sessions and the opportunity to perform practice interviews. However, field control procedures, and particularly interviewer supervision, have varied across countries. In some cases, fieldwork was highly controlled but in others the control was very limited. Some but not all of the countries employed a check on 10% of cases to confirm that an interview had taken place and to find out what respondents thought about the interviewer and the study.

Germany and Sweden used small incentives to gain respondents' cooperation. This practice was a deviation from the design specifications. In order to increase the response rate, Germany and Switzerland told respondents that one of the aims of the survey was to improve the quality of newspapers, instructions and forms written for the public. For the same purpose, Poland presented the survey as a test rather than a real study.

The information we have regarding the data collection procedures is based on questions that we asked the different organizations and the survey managers involved. It has been difficult to get a good understanding of all the data collection aspects involved, but again the need for standardization should be emphasized. Aspects that should be standardized include interviewer training, presentation of the study to sample persons, field work control and supervision as well as measures for reducing nonresponse rates. In addition standardized procedures need to be strictly applied in administering the exercise booklets, including the possibility of skipping items, the use of calculators, assistance from other persons, and writing responses on a separate sheet. Furthermore, issues of respondent burden need to be considered in relation to standardized procedures. If a country makes extra demands on the survey respondents, for instance by asking them to also complete a battery of its own test items, the increased response burden may affect the amount of effort respondents make in answering the IALS items. Finally, standardized procedures need to be followed for handling persons who speak only a foreign language, who are too ill, and who complete only part of an interview.

9. Data Processing and Data Entry

Each country was responsible for coding industry, occupation, and education to the ISIC, ISOC, and ISCED standards at the finest level possible. Outgoing quality levels at the one digit level should exceed 95%. This coding task is a formidable one. First, there is a problem getting consistent coding when it is conducted in a decentralized fashion. Coding in an international survey using different types of survey organization is an extreme form of decentralization. Second, the 95% quality requirement is a tough one; error rates in the range of 5-15% are not uncommon. We have not been able to collect much information on this issue. Agencies like Statistics Canada and the U.S. Census Bureau have reliable facilities and sufficient know-how to deal with this task. Some of the other organizations involved probably do not know very much about the error problems associated with this kind of complex coding.

Each country was responsible for coding the exercise booklets, and this task can also present difficulties. The rules establishing which answers are "right" and "wrong" were not always easy for the coders to understand. Sometimes, because of a special situation in a particular country, borderline answers were treated as "right", and in other cases not.

Each country was instructed to perform 100% verification of the background questionnaire data entry and the entry of strings of test scores. It turns out that not all countries performed the 100% verification. Their efforts range from no verification, through sampling verification, to 100% verification.

10. Weighting

Weighting of responding sample units is performed to compensate for unequal selection probabilities, for nonresponse, and for noncoverage. The weights are often developed in three stages: first, the construction of base weights that are the inverses of selection probabilities; second, the adjustment of the base weights to compensate for nonresponse; and third, a further adjustment of the weights to make the weighted sample distributions for certain variables conform to the distributions for these variables available from external benchmark sources.

The countries had varying ambitions in carrying out the weighting task. Some countries computed the base weights incorrectly. We were not able to determine the effect of this error on the survey estimates. We recommend that this issue be examined in subsequent evaluations of the surveys.

In view of the high nonresponse rates in these surveys, we consider it particularly important that the base weights be adjusted as effectively as possible to compensate for nonresponse and noncoverage. In general, we believe that more attention should have been given to this issue. Since literacy is highly related to education, post-stratification to a current population education distribution is particularly recommended, provided that education is measured in the same way by the survey and by the external source. As a general rule, when consideration is given to adjusting sample weights to make the weighted sample distributions conform to benchmark distributions from external sources, attention should be paid to ensuring that the variables are measured in the same way by the survey and the external sources, and that the external sources provide up-to-date distributions for the variables.

The primary purpose of weighting adjustments is to reduce the bias in the survey estimates. However, large variation in weights can seriously inflate the variances of the estimates. In some cases, the variation in weights was as much as 15-fold. In such cases, consideration should be given to trimming the weights, thereby accepting some bias but reducing the variances of the estimates. Major adjustments to the base weights give rise to a serious concern about the representativeness of the sample, for they indicate that some groups of the population are substantially underrepresented in the sample. The adjustments serve to correct that underrepresentation, but they are based on the questionable assumption that the sample respondents in a group have similar survey characteristics to those missed by the survey.

11. Conclusions

Our review of the data quality of the first round of the IALS has been carried out under considerable time pressure. Using the questionnaire we developed, we have collected a good deal of information about the survey procedures used in each of the eight countries. This information has in many cases led us to the desire for more data and for some empirical analyses of the surveys. However, in general, both time and resources precluded us from further investigation. Thus while we have identified a number of aspects of the surveys that give us cause for concern, we have not been able to make any real assessments about the effects of the various error sources.

As the result of our work, we have two types of recommendation to make. One type concerns the first round of the IALS and the second concerns the future rounds. With regard to the first round of the survey, we have the following three recommendations:

1. The descriptions of the survey procedures employed in the various countries that we have obtained show that there are problems in most design components across countries, and especially with regard to sample design and nonresponse. In our view, these problems threaten the validity of any comparisons of literacy levels across countries. We therefore recommend that no tables ranking countries by literacy levels be published.
2. Although we oppose the ranking of countries, we believe that the first round of the IALS has collected valuable data that should be published. We recommend that the publications focus on the correlates of literacy in the different countries and on the comparison of these correlates across countries. We believe that these forms of analysis may be less affected by the methodological problems we have noted. Nevertheless, the report should contain cautions to warn the reader about the possible effects that these methodological problems might have.
3. Our limited study has identified a number of areas where further methodological research is needed to assess the quality of the IALS data. We recommend that, where possible, methodological analyses be conducted to study, for example, the characteristics of nonrespondents, the comparison of survey estimates to benchmark data (before reweighting), and the effect of alternative weighting schemes. The overall aim of this research is to provide empirical evidence on the quality of the IALS data.

We have two major recommendations for future rounds of the IALS survey:

A. In addition to the use of a carefully-constructed standardized measuring instrument, there is also need for a standardization of survey procedures across the countries. Best practices need to be established for all the components of the survey, and these practices need to be followed in the participating countries. The achievement of this standardization using best practices requires more than the specification of the practices to be followed. Careful monitoring is also required. The implementation of this recommendation clearly has cost implications. We are certain, however, that the investment will pay off.

B. In commissioning survey organizations to conduct the IALS survey, a major consideration should be the capability of those organizations to conduct the survey to the standards required. The organizations should have strong records in all aspects of survey research, including probability sampling, data collection, data processing and weighting.

In conclusion, we should note that we fully recognize the considerable difficulties in achieving standardization using best practices in international surveys. We believe, however, that for such surveys to achieve their full potential, this is what is needed. The first round of the IALS has produced some valuable results on literacy in the eight participating countries. We believe that if the above recommendations are followed, future rounds of the survey will make even greater contributions to the international study of literacy.

Appendix B:

IALS Background Questionnaire

International Adult Literacy Survey
Background Questionnaire

Confidential once Completed
Collected under the authority of the Statistics Act, Revised Statutes of Canada, 1985 Chapter S19.
Version francaise disponible

1. Language of Interview

X English __French

2. Total Number of Calls: __ __

3. Interviewer name:________________________________

4. Was any assistance provided by a third party for the completion of the background questionnaire (BQ)?
__Yes
__No

5. Final Status MAIN BQ CORE
__ __ __

Interviewer: *Please provide a detailed reason for non-response in the area provided on page 23 of this questionnaire.*

Record of Calls and Appointments

	Date	Start Time	Finish Time	Notes
1				
2				
3				
4				
5				

INTRODUCTION

Hello, this is ...from Statistics Canada.

Statistics Canada is conducting a national literacy survey in cooperation with the department of Human Resources Development and the National Literacy Secretariat.

Results from the survey will be used to plan programs suited to the needs of Canadians. It will also be used to compare Canadian needs with those of other countries who are conducting a similar study. Your voluntary participation is needed if the results are to be accurate. Your answers will be kept confidential and combined with those of others to produce statistics.

This survey requires that I meet with you in person and should take about 45 to 75 minutes. Would it be possible to meet on...at...?

Would you prefer to be interviewed in English or French

Do you still live at...? (Confirm address on label)

Part 1 - Background Questionnaire

6. INTERVIEWER:
Record start time of BQ: __ __ : __ __

SECTION A. GENERAL INFORMATION

A1. **First I'd like to ask you a few questions about your background, your education, the languages you speak and the jobs you may have held in the past 12 months.**

Were you born in Canada?

__Yes —> *Go to Q. A6*
__No

A2. **In what country were you born?**

__United Kingdom
__Italy
__United States
__Poland
__Germany
__Other - *Specify*

__India
__Portugal
__Peoples Rep. of China
__Hong Kong
__Netherlands

____
____

A3. **In what year did you first immigrate to Canada?**

19__ __ ***OR***

__Canadian Citizen by birth —> *Go to Q. A7*

A4. **In total how many years have you lived in Canada?**

__ __ years

A5. **Before you first immigrated to Canada, what was the highest level of schooling you had completed?** *(Mark one only)*

__No Education
__Did not complete Primary
__Completed Primary
__Some secondary
__Completed secondary (Vocation or Technical)
__Completed secondary (General or Academic)
__Completed non-university Post-Secondary
__ Completed university
__Completed university Post-Graduate
__Education not definable by level

—>> *Go To Q. A7.*

A6. **In what Province were you born?**

__Newfoundland
__Prince Edward Island
__Nova Scotia
__New Brunswick
__Quebec
__Ontario
__Manitoba
__Saskatchewan
__Alberta
__British Columbia
__North West Territories
__Yukon

A7. **During your lifetime, how many years of formal education have you completed beginning with grade one and not counting repeated years at the same level ?**

__ __ If **00** —> (No education) *Go To Q. B1*

A8. **What is the highest level of schooling you have ever completed?** *(Mark one only)*

__Less than Primary —> *Go to Q. A12*
__Completed Primary —> *Go to Q. A12*
__Some secondary —> *Go to Q. A11*
__Completed secondary
__Completed non-university Post-Secondary
__Completed university
__Completed university Post-Graduate
__Education not definable by level

A9. **Which of the following best describes your secondary program? Was it...** *(Mark one only)*

__**an academic / college preparatory type program?**
__**a business (commercial) or trade / vocational type program?**
__**a high school equivalency program?**
__Don't know / not applicable

A10. **In what province or country did you complete your secondary education?**

__Newfoundland
__Prince Edward Island
__Nova Scotia
__New Brunswick
__Quebec
__Ontario
__Manitoba
__Saskatchewan
__Alberta
__British Columbia
__North West Territories
__Yukon
__Other-Specify

__United Kingdom
__India
__Italy
__Portugal
__United States
__Peoples Rep. of China
__Poland
__Hong Kong
__Germany
__Netherlands

____

____

—>> *Go to Q. B1*

A11. **Which of the following best describes your secondary program? Was it...** *(Mark one only)*

__**an academic / college preparatory type program?**
__**a business (commercial) or trade / vocational type program?**
__**a high school equivalency program?**
__Don't know / not applicable

A12. **What was the main reason you stopped your schooling when you did?** *(Mark one only)*

__Still in school
__Had enough education
__Had to work / financial reasons
__Wanted to work / wanted to learn a trade
__Family reasons (help family business, illness at home, marriage, pregnancy, etc.)
__Did not like school / boredom
__Did not do well in school
__Personal Illness or disability
__School not available / not accessible
__To join the military
__Don't know
__Other - specify

____

____

SECTION B. LINGUISTIC INFORMATION

B1. **What language did you FIRST speak as a child?**

(Accept multiple responses only if languages were spoken EQUALLY)

__English —> *Go to Q. B12*
__French
__Italian
__Chinese
__German
__Portuguese
__Polish
__Ukrainaian
__Spanish
__Dutch
__Punjabi
__Greek

__REFUSED

__Other 1 - Specify

____

__Other 2 - Specify

____

B2. **How would you rate your current ability to speak that language?** *(read categories)*

(INTERVIEWER: Ask for each language specified in Q.B1)

	First Language	Second Language
INTERVIEWER: *Insert Code from Q.B1*	— —	— —
Cannot speak that language	—	—
Poor	—	—
Fair	—	—
Good	—	—
Very Good	—	—

B3. **How would you rate your current ability to understand that language when it is spoken to you?** *(read categories)*

(INTERVIEWER: Ask for each language specified in Q.B1)

	First Language	Second Language
INTERVIEWER: *Insert Code from Q.B1*	— —	— —
Cannot understand that language	—	—
Poor	—	—
Fair	—	—
Good	—	—
Very Good	—	—

B4. **How would you rate your current reading skills in that language?** *(read categories)*

(INTERVIEWER: Ask for each language specified in Q.B1)

	First Language	Second Language
INTERVIEWER: *Insert Code from Q.B1*	— —	— —
Cannot read in that language	—	—
Poor	—	—
Fair	—	—
Good	—	—
Very Good	—	—

B5. **How would you rate your current writing skills in that language?** *(read categories)*

(INTERVIEWER: Ask for each language specified in Q.B1)

INTERVIEWER:	First Language	Second Language
Insert Code from Q.B1	__ __	__ __
Cannot write in that language	__	__
Poor	__	__
Fair	__	__
Good	__	__
Very Good	__	__

B6. **How old were you when you first started to learn English?**
(INTERVIEWER:Insert lowest age mentioned)

__ __years old

__Does not speak English

B7. **When you were growing up, what language or languages were usually spoken in your home?**

(INTERVIEWER: Accept multiple responses only if languages were spoken EQUALLY)

__English
__French
__Italian
__Chinese
__German
__Portuguese
__Polish
__Ukrainian
__Spanish
__Dutch
__Punjabi
__Greek

__REFUSED
__Other 1 - Specify

____

__Other 2 - Specify

____

B8. **What language did you first learn to read and write?**

__English
__French
__Italian
__Chinese
__German
__Portuguese

__Polish
__Ukrainian
__Spanish
__Dutch
__Punjabi
__Greek

__REFUSED

__Other 1 - Specify

____

__Other 2 - Specify

____

B9. **Have you ever taken a course to learn English?**

__Yes __No

B10. **How well do you understand English when it is spoken to you?**

__Poorly
__Fairly well
__Well
__Very well
__Cannot understand English

B11. **How well can you speak English?**

__Poorly —>
__Fairly well —>
__Well —> ***all answers Go to Q.B13***
__Very well —>
__Cannot speak English —>

B12. **Since leaving school, have you ever taken a course to upgrade your reading or writing skills in English?**

__Yes __No __Still in school

B13. **What languages (including English) do you speak well enough to conduct a conversation?** *(Mark all that apply)*

—>>> *Interviewer: if only one language, Go to Q.B18*

__English
__French
__Italian
__Chinese
__German
__Portuguese
__Other - Specify
__Polish
__Ukrainian
__Spanish
__Dutch
__Punjabi
__Greek

____

____

B14. **What language do you speak most often at home?** *(Mark one only)*

__English
__French
__Italian
__Chinese
__German
__Portuguese
__Other - Specify
__Polish
__Ukrainian
__Spanish
__Dutch
__Punjabi
__Greek

____

____

B15. **What language do you speak most often at work or school?** *(Mark one only)*

__not applicable

__English
__French
__Italian
__Chinese
__German
__Portuguese
__Other - Specify
__Polish
__Ukrainian
__Spanish
__Dutch
__Punjabi
__Greek

____

____

B16. **What language do you speak most often during leisure activities?** *(Mark one only)*

__English
__French
__Italian
__Chinese
__German
__Portuguese
__Other - Specify

__Polish
__Ukrainian
__Spanish
__Dutch
__Punjabi
__Greek

____

____

B17. **In which language can you express yourself most easily?** *(Mark one only)*

__English
__French
__Italian
__Chinese
__German
__Portuguese
__Other - Specify

__Polish
__Ukrainian
__Spanish
__Dutch
__Punjabi
__Greek

____

____

B18. **To which ethnic or cultural group did your ancestors belong?** *(Mark all that apply)*

__English
__French
__Italian
__Ukrainian
__German
__Irish
__Metis
__North American Indian
__Other - Specify

__Chinese
__Dutch
__Scottish
__Jewish
__Polish
__Portuguese
__Inuit
__Canadian

____

____

SECTION C. PARENTAL INFORMATION

C1. **The next few questions are about your mother** (female guardian). **Can you answer some questions about her?**

__Yes
__No —> *Go to Q. C7*

C2. **Was your mother** (female guardian) **born in Canada?**

__Yes —> *Go to Q. C5*
__No
__Don't know —> *Go to Q. C5*

C3. **Did your mother** (female guardian) **immigrate to Canada?**

__Yes
__No —> *Go to Q. C5*

C4. **Was your mother** (female guardian) **under the age of 16 when she immigrated to Canada?**

__Yes
__No
__Don't know

C5. **What was the highest level of schooling that your mother** (female guardian) **ever completed?** *(Mark one only)*

__No Education
__Did not complete Primary
__Completed Primary
__Some secondary
__Completed secondary (Vocation or Technical)
__Completed secondary (General or Academic)
__Completed non-university Post-Secondary
__Completed university
__Completed university Post-Graduate
__Education not definable by level
__Don't know

C6. **Did/does your mother (female guardian) work at a job or business?**

__Yes
__No
__Don't know

C7. **The next few questions are about your father** (male guardian). **Can you answer some questions about him?**

__Yes
__No —> *Go to Q. D1*

C8. **Was your father** (male guardian) **born in Canada?**

__Yes —> *Go to Q. C11*
__No
__Don't know —> *Go to Q. C11*

C9. **Did your father** (male guardian) **immigrate to Canada?**

__Yes
__No —> *Go to Q. C11*

C10. **Was your father** (male guardian) **under the age of 16 when he immigrated to Canada?**

__Yes
__No
__Don't know

C11. **What was the highest level of schooling that your father** (male guardian) **ever completed?** *(Mark one only)*

__No Education
__Did not complete Primary
__Completed Primary
__Some secondary
__Completed secondary (Vocation or Technical)

__Completed secondary (General or Academic)
__Completed non-university Post-Secondary
__Completed university
__Completed university Post-Graduate
__Education not definable by level
__Don't know

C12. **For whom did / does your father** (male guardian) **work?** *(Probe for main) (Name of business government dept. or agency, or person)*

____
____

__Don't Know

__Never worked --> *Go to QD1*

C13. **What kind of business, industry or service is / was this?** *(Give full description: e.g., federal government, canning industry, forest services.)*

____
____

__Don't Know

C14. **What kind of work was / is your father** (male guardian) **doing?** *(Give full description: e.g., office* clerk, factory worker, forest technician.)

____
____

__Don't Know

SECTION D. LABOUR FORCE INFORMATION

D1. **I would now like to talk about your employment status. What is your current work situation? Are you...**
(mark one only)

__**Employed** —> *Go to Q. D4*
__**Retired**
__**Unemployed / looking for work**
__**Student (including Work Programs)**
__**Homemaker**
__Other

D2. **Did you work at a job or business at any time in the past 12 months (regardless of the number of hours per week)?**

__Yes —> *Go to Q. D4*
__No

D3. **When did you last work at a job or business?**

19__ __ —> *Go to Q. D19*

__Never Worked —> *Go to Q. D19*

D4. **How many different employers have you had in the past 12 months?**
__ __ employer(s)

D5. **Did you work mostly full-time (that is, more than 30 hours per week) or part-time (that is, less than 30 hours per week)?**

__Full -Time —> Go to Q. D7
__Part-Time

D6. **Why did you work part-time?** *(Mark one only)*

__Own illness or disability
__Child care responsibilities
__Other personal or family responsibilities
__Going to school or taking training
__Could only find part-time work
__Did not want to work full-time
__Retired
__Other - Specify

____
____

D7. **For whom did you work the most hours in the past 12 months?** (*name of business, government department, or person)*

____
____

D8. **What kind of business, industry or service was this?** *(Give full description, e.g. fish canning plant, automobile manufacturing plant, municipal government)*

____
____

D9. **What kind of work were you doing at this job?** *(Give full description or occupational title, e.g. office clerk, machine operator, computer programmer)*

____
____

D10. **In total, about how many persons are employed by this business at all locations in Canada?**

__Less than 20?
__20 to 99?
__100 to 199?
__200 to 499?
__500 or over?
__Don't Know

D11. **What was your status at this job? Was it as an...**

__ **Employee without supervisory responsibilities**

__ **Employee with limited supervisory or management responsibilities (5 persons or less)**

__ **Employee with more extensive supervisory or management responsibilities (more than 5 persons)**

__ **Self-employed without employees**

__ **Self employed with employees**

__ **Family worker (unpaid)**

D12. **What type of job was this? Was or is this job a...**

__**Permanent job or work contract of unlimited duration**
__**Temporary job or work contract of limited duration**

D13. **How many hours per week did you usually work at this job?**

__ __Hours

D14. **During the past 12 months, how many weeks did you work (at all jobs including time off for vacation, maternity leave, illness, strikes and lockouts)?**

__ __ weeks —> *If 52 Go To Q.E1*

D15. **During the past 12 months, in the weeks when you were without work, did you want to work?**

__Yes —> *Go To Q. D17*
__No

D16. **Why did you not want to work?** *(Mark one only)*

__Own illness or disability
__Child care responsibilities
__Other personal or family responsibilities
__Going to school or taking training
__Retired
__Not interested in working
__Other - Specify

____

____

—>> ***all answers Go to Q. E1***

D17. **During the past 12 months, for how many weeks were you without work and NOT looking for work?**

__ __weeks —> If **00** *Go to Q. E1*

D18. **What is the main reason why you did not look for work during these weeks?** *(Mark only one)*

__Own illness or disability
__Child care responsibilities
__Other personal or family responsibilities
__Awaiting recall from a temporary layoff
__Waiting for a job to start
__Did not have the skills or experience for available jobs
__Too old to work / Retired
__Other - Specify

____

____

—>> ***all answers Go to Q. E1***

D19. **During the past 12 months, in the weeks when you were not working at a job or business, did you want to work?**

__Yes —> *Go to Q. D21*
__No

D20. **Why did you not want to work?**

__Own illness or disability
__Child care responsibilities
__Other personal or family responsibilities
__Going to school or taking training
__Retired
__Not interested in working
__Homemaker
__Other - Specify

____

____

—>> *all answers Go to Q. F1*

D21. **During the past 12 months, for how many weeks were you without work and NOT looking for work?**

__ __weeks —> If **00** *Go to Q. F1*

D22. **What is the main reason why you did not look for work during these weeks?** *(Mark one only)*

__Own illness or disability
__Child care responsibilities
__Other personal or family responsibilities
__Awaiting recall from a temporary layoff
__Waiting for a job to start
__Did not have the skills or experience for available jobs
__Too old to work / Retired
__Other - Specify

____

____

—>> *all answers Go to Q. F1*

SECTION E READING AND WRITING AT WORK AND LOOKING FOR WORK

E1. **The following questions refer to the job at which you worked the most hours in the last 12 months**

How often (do/did) you read or use information from each of the following as part of your main job? Would you say every day, a few times a week, once a week, less than once a week, rarely or never?

		Every day	A few times a week	Once a week	Less than once a week	Rarely or never
A.	Letters or memos	__	__	__	__	__
B.	Reports, articles, magazines or journals	__	__	__	__	__
C.	Manuals or reference books, including catalogues.	__	__	__	__	__
D.	Diagrams or schematics	__	__	__	__	__
E.	Bills, invoices, spreadsheets or budget tables	__	__	__	__	__
F.	Material written in a language other than English	__	__	__	__	__
G.	Directions or instructions for medicines, recipes, or other products	__	__	__	__	__

E2. **How often (do/did) you write or fill out each of the following as part of your main job? Would you say every day, a few times a week, once a week, less than once a week, rarely or never?**

		Every day	A few times a week	Once a week	Less than once a week	Rarely or never
A.	Letters or memos	__	__	__	__	__
B.	Forms or things such as bills, invoices, or budgets	__	__	__	__	__
C.	Reports or articles	__	__	__	__	__
D.	Estimates or technical specifications	__	__	__	__	__

E3. **In your main job, how often do you use arithmetic or mathematics (that is, adding, subtracting, multiplying or dividing) to:**

A) measure or estimate the size or weight of objects?

Every day	__
A Few times a week	__
Once a week	__
Less than once a week	__
Never	__

B) Calculate prices, costs or budgets?

Every day.	__
A Few times a week	__
Once a week	__
Less than once a week	__
Never	__

E4. **How would you rate your reading skills in English for your main job?**

__**Excellent**
__**Good**
__**Moderate**
__**Poor**
__No opinion/Not Applicable

E5. **To what extent are your reading skills in English limiting your job opportunities - for example, advancement or getting another job?**

__**Greatly limiting**
__**Somewhat limiting**
__**Not at all limiting**

E6. **How would you rate your writing skills in English for your main job?**

__ **Excellent**
__**Good**
__**Moderate**
__**Poor**
__No opinion/not applicable

E7. **To what extent are your writing skills in English limiting your job opportunities - for example, advancement or getting another job?**

__**Greatly limiting**
__**Somewhat limiting**
__**Not at all limiting**

E8. **How would you rate your mathematical skills for your main job?**

__**Excellent**
__**Good**
__**Moderate**
__**Poor**
__No opinion/not applicable

E9. **To what extent are your mathematical skills limiting your job opportunities - for example, advancement or getting another job?**

__Greatly limiting
__Somewhat limiting
__Not at all limiting

SECTION F. ADULT EDUCATION

F1. **The following questions will deal with any education or training which you may have taken in the past 12 months.**

During the past 12 months, that is, since August 1993, did you receive any training or education including courses, private lessons, correspondence courses, workshops, on- the-job training, apprenticeship training, arts, crafts, recreation courses or any other training or education?

__Yes

__No —> *Go To Q.F15*

F2. **In total, how many courses did you take in the past 12 months?**

— —

	FIRST MENTION	SECOND MENTION	THIRD MENTION
F3. **What were the names (titles) of these courses or the program associated with these courses?**	____________ ____________ ____________ ____________	____________ ____________ ____________ ____________	____________ ____________ ____________ ____________

INTERVIEWER: Insert the names of the three most recent courses/programs in the space provided and prompt for multiple responses.

Note: A program is a collection of courses which leads to a specific degree, diploma or certificate. If the respondent lists many courses that all begin at the same time, probe for a program name.

{ASK THE REMAINING QUESTIONS IN SECTION F UNTIL QUESTION F15 FOR ALL THREE COURSE/PROGRAMS MENTIONED IN QUESTION F3}

F4. **Now I'd like to ask you about...***[insert name of course or program name]*
Was this training or education financially supported by...
(Read categories)

	Yes	No
a) yourself or your family?...	___	___
b) an employer?...	___	___
c) the government?...	___	___
d) a union or professional organization?...	___	___
e) anyone else?...	___	___
f) no fees...	___	___
g) don't know...	___	___

F5. **Were you taking this training or education towards...**
(Read categories)
(Mark one only)

a university degree/diploma/certificate? ___
a college diploma/certificate? ___
a trade-vocational diploma/certificate? ___
an apprenticeship certificate? ___
an elementary or secondary school diploma? ___
professional or career upgrading? ___
other.... ___

{REMEMBER: ASK THE REMAINING QUESTIONS IN SECTION F UNTIL QUESTION F15 FOR ALL THREE COURSE/PROGRAMS MENTIONED IN QUESTION F3}

F6. **Was this training or education given by...**
(Read categories)

	Yes	No
a) **a university or higher education establishment?...**	___	___
b) **a further education college?...**	___	___
c) **a commercial organization (for example, a private training provider)?...**	___	___
d) **a producer or supplier of equipment?...**	___	___
e) **a non-profit organization such as an employer association, voluntary organization or a trade union?...**	___	___
f) **an employer or a parent company?...**	___	___
g) other...	___	___

F7. **Where did you take this training or education?**
(Mark one only)

___ **Elementary or High School**
___ **College Campus**
___ **University Campus**
___ **Business or Commercial School**
___ **Work**
___ **Training centre**
___ **Conference centre or hotel**
___ **Home**
___ **Community centre or sports facility**
___ **Elsewhere**

F8. **For how many weeks did this training or education last?** _____ **weeks**

{REMEMBER: ASK THE REMAINING QUESTIONS IN SECTION F UNTIL QUESTION F15 FOR ALL THREE COURSE/PROGRAMS MENTIONED I N QUESTION F3}

F9. **On average, how many days per week was it?** _____ days

F10. **On average, how many hours per day was it?** _____ hours

F11. **What was the main reason you took this training or education? Was it for...**

(Read categories)

(Mark one only)

career/job related purposes?...	___	*--> Go to Q. F12*
personal interest?...	___	*--> Go to Q. F13*
other...	___	*--> Go to Q. F13*

F12. **To what extent are you using the skills or knowledge acquired in this training or education at work?**

(Read categories)

To a great extent...	___
Somewhat...	___
Very little...	___
Not at all...	___
Not applicable...	___

{REMEMBER: ASK THE REMAINING QUESTIONS IN SECTION F UNTIL QUESTION F15 FOR ALL THREE COURSE/PROGRAMS MENTIONED I N QUESTION F3}

F13. **Who suggested you take this training or education?**
(Read categories)

		Yes	No
a)	**You did...**	___	___
b)	**Your friends or family...**	___	___
c)	**Your employer...**	___	___
d)	**Other employees**	___	___
e)	**Part of a Collective Agreement...**	___	___
f)	**Your Union or trade association...**	___	___
g)	**Legal or professional requirement...**	___	___
h)	**Social Services or Employment Centre...**	___	___
i)	Other...	___	___
j)	Don't know...	___	___

F14. **Was this training or education provided through...**
(Read categories)

		Yes	No
a)	**classroom instruction, seminars or workshop?...**	___	___
b)	**educational software?...**	___	___
c)	**radio or TV broadcasting?...**	___	___
d)	**audio/video cassettes, tapes or disks?...**	___	___
e)	**reading materials?...**	___	___
f)	**on-the-job training?...**	___	___
g)	other methods?...	___	___

F15. **Since August 1993, was there any training or education that you WANTED to take for career or job-related reasons but did not?**

__Yes

__No —> *Go to Q.F17*

F16. **What were the reasons you did not take this training or education?** *(Mark all that apply)*

__Too busy/lack of time
__Too busy at work
__Course not offered
__Family responsibilities
__Too expensive/no money
__Lack of qualifications
__Lack of employer support
__Course offered at inconvenient time
__Language reasons
__Health reasons
__Other

F17. **Since August 1993, was there any other training that you WANTED to take but did not, such as hobby, recreational or interest courses?**

__Yes

__No —> *Go to Q.G1*

F18. **What were the reasons you did not take this training or education? (Mark all that apply)**

__Too busy/lack of time
__Too busy at work
__Course not offered
__Family responsibilities
__Too expensive/no money
__Lack of qualifications
__Lack of employer support
__Course offered at inconvenient time
__Language reasons
__Health reasons
__Other

SECTION G. READING AND WRITING GENERAL

G1. **The next few questions deal with reading and writing in your daily life excluding work or school.**

I am going to read you a list of activities. Please tell me if you do each of them daily, weekly, every month, several times a year or never? How often do you...

		Daily	Weekly	Monthly	Several times a year	Never
a)	use a public library?...	___	___	___	___	___
b)	attend a movie, play or concert?...	___	___	___	___	___
c)	attend or take part in a sporting event?...	___	___	___	___	___
d)	write letters or anything else that is more than one page in length?...	___	___	___	___	___
e)	participate in volunteer or community organizations?...	___	___	___	___	___
f)	read newspapers or magazines?...	___	___	___	___	___
g)	read books?...	___	___	___	___	___
h)	listen to radio, records, tapes, cassettes or compact discs?...	___	___	___	___	___

G2. **Do you ever do any of these activities in a language other than French or English?**

__Yes

__No —> *Go to Q.G4*

G3. **Which of the following activities have you ever done in a language other than French or English?**

		Yes	No
A.	**use a public library. ....**	__	__
B.	**attend a movie, play or concert. ....**	__	__
C.	**attend or take part in a sporting event. ....**	__	__
D.	**write letters or anything else that is more than one page in length ....**	__	__
E.	**participate in volunteer or community organizations ....**	__	__
F.	**reading newspapers or magazines ....**	__	__
G.	**read books. ....**	__	__
H.	**listening to radio, records, tapes, cassettes or compact discs. ....**	__	__

G4. **How much time do you usually spend each day watching television or videos?**

__Not on a daily basis
__1 hour or less per day
__1 to 2 hours per day
__more than 2 hours but less than five
__5 or more hours per day

__Do not have a television or videos —> *Go to Q.G6*

G5. **Do you ever watch television or videos in a language other than French or English?**

__Yes __No

G6. **Which of the following materials do you currently have in your home?**

	Yes	No
daily newspapers.	__	__
weekly newspaper / magazines	__	__
more than 25 books	__	__
a (multi-volume) encyclopedia	__	__
a dictionary	__	__

G7. **How often (do/did) you read or use information from each of the following as part of your daily life? Would you say every day, a few times a week, once a week, less than once a week, rarely or never?**

	Every day	A few times a week	Once a week	Less than once a week	Rarely or never
A. **Letters or memos**					
B. **Reports, articles, magazines or journals**					
C. **Manuals or reference books, including catalogues**					
D. **Diagrams or schematics..**					
E. **Bills, invoices, spreadsheets or budget tables**					
F. **Material written in a language other than English**					
G. **Directions or instructions for medicines, recipes, or other products**					

G8. **I am now going to read you a list of different parts of a newspaper. Please tell me which parts you generally read when looking at a newspaper.** *(Mark all that apply)*

	Yes	No
classified ads	__	__
other advertisements	__	__
national/international news ..	__	__
regional or local news	__	__
sports	__	__
home, fashion, or health	__	__
editorial page	__	__
financial news or stock listings	__	__
comics	__	__
TV listings	__	__
movie or concert listings	__	__
book, movie or art reviews ...	__	__
horoscope	__	__
advice column	__	__
other - *specify.*	__	__

__do not read the newspaper

G9. **Would you say you follow what's going on in current events, government and public affairs...**

__most of the time

__some of the time

__only now and then

__hardly at all

G10. **I would like to know how you usually get information about current events, public affairs, and the government. How much information do you get from...**

		A lot	Some	Very little	None
A.	newspapers	__	__	__	__
B.	magazines	__	__	__	__
C.	radio	__	__	__	__
D.	television	__	__	__	__
E.	family members, friends or co-workers	__	__	__	__

G11. **Sometimes people need help from family members or friends to read and write in English. How often do you need help from others with...**

	Often	**Sometimes**	**Never**
a) **Reading newspaper articles**	__	__	__
b) **Reading information from government agencies, businesses or other institutions**	__	__	__
c) **Filling out forms such as applications or bank deposit slips.**	__	__	__
d) **Reading instructions such as on a medicine bottle**	__	__	__
e) **Reading instructions on "packaged" goods in stores or supermarkets**	__	__	__
f) **Doing basic arithmetic, that is, adding, subtracting, multiplying and dividing.**	__	__	__
g) **Writing notes and letters.**	__	__	__

G12. **How would you rate your reading skills in English needed in daily life?**

__**Excellent**
__**Good**
__**Moderate**
__**Poor**
__No opinion

G13. **How would you rate your writing skills in English needed in daily life?**

__Excellent
__Good
__Moderate
__Poor
__No opinion

G14. **How would you rate your mathematical skills needed in daily life?**

__Excellent
__Good
__Moderate
__Poor
__No opinion

G15. **All things considered, how satisfied are you with your reading and writing skills in English? Are you...**

__Very Satisfied
__Somewhat satisfied
__Somewhat dissatisfied
__Very dissatisfied
__No opinion

	16. **Did you ever have...**	17. **Did you have this problem while you were in primary or secondary school?**	G18. **Do you have this problem now?**
	No/Yes If yes-->	**No/Yes** If yes -->	**No/Yes**
a)	**eye/ visual trouble of the kind that is not corrected by glasses? ....**		
b)	**hearing problems? ....**		
c)	**a speech disability? ....**		
d)	**a learning disability? ....**		
e)	**any other disability or health problem of six months or more? ....**		

SECTION H. FAMILY LITERACY

H1. **Are you the parent or guardian of any children aged 6 to 18 that are presently living with you?**

__Yes
__No —> *Go to Q. J1*

H2. **What is the age of your youngest child between 6 to 18 years of age?**

__ __Years Old

H3. **What is the highest grade of schooling that this child has completed?**

Elementary ——> Grade__ __
Secondary ——> Grade__ __
Post-Secondary ——> Year__ __
Trade/Vocational ——> Year__ __

__No Schooling

H4. **How often would you say this child reads for pleasure? Would you say...**

__Every day
__A few times a week
__Several times a month
__A few times a month
__Once a month or less
__Never
__Don't know
__Not Applicable

H5. **When this child reads, where does he/she get books?** *(Mark all that apply)*

__parent buys
__parent borrows from a friend
__child buys
__child borrows from a friend
__public library
__school library
__gifts
__from brothers/sisters
__other
__don't know

H6. **Given this child's age, how satisfied are you with the way he/she reads? Would you say you are...**

__**Very satisfied**
__**Somewhat satisfied**
__**Somewhat dissatisfied**
__**Very dissatisfied**
__No opinion

The next few questions will deal with your entire household.

H7. **Could you please tell me if each of the following statements are true or false of your household?**

	True	**False**	**Don't know**
A. **There is a variety of books in your home**	__	__	__
B. **There is a variety of magazines and other reading material in your home**	__	__	__
C. **Your children often see you or your spouse reading**	__	__	__
D. **Your children learned to read before grade one**	__	__	__
E. **Your children have a certain amount of time set aside each day for reading at home**	__	__	__
F. **Your children are limited in the amount of time you allow them to watch TV**	__	__	__
G. **Your children often choose the books they read**	__	__	__
H. **Your children have their own books and a place to keep them**	__	__	__

H8. **I would like to read you a list of some different things which may help parents in helping their children to become good readers. For each one, I would like you to tell me whether it is something that you would find very useful, somewhat useful, not very useful or not at all useful.**

	Very	Some-what	Not very	Not at all
a) **Reading lists supplied to parents by schools and other educational experts...**	__	__	__	__
b) **Materials such as games, activities and books being supplied to parents by schools or other educational experts that would help parents encourage their children to read...**	__	__	__	__
c) **A close parent-teacher relationship...**	__	__	__	__
d) **Schools providing parents with help in understanding assessments of a child's reading abilities...**	__	__	__	__
e) **Access for children to books, through either public or school libraries...**	__	__	__	__

SECTION J. HOUSEHOLD INFORMATION

J1. **Finally, I would like to get a bit of general household information.**

From which of the following sources did you receive income in 1993?

		Yes	No
a)	**Income from wages, salary or self-employment ....**	__	__
b)	**Income from government, such as Family allowance, Unemployment Insurance or Social Assistance ....**	__	__
c)	**Canada or Quebec Pension Plan, or old age pension ....**	__	__
d)	**Income from interest, dividends, investments or private pensions.................**	__	__
e)	**Income from any other sources, such as alimony, scholarships, etc ....**	__	__

J2. **What is the best estimate of your personal income in 1993 from all sources, including those just mentioned?**

__ __ __ __ __ __ __ __.00

__No Income —> *Go to Q. J4*

__Don't know

J3. **What is the best estimate of your personal income from only wages, salary or self-employment in 1993?**

__ __ __ __ __ __ __ __.00

__No Income

__Don't know

J4. **Including yourself, how many people live in this household?**

__ __ —> if 01 *Go to Q.7*

J5. **What is the best estimate of the total income of all household members (including youself) from all sources in 1993?**

__ __ __ __ __ __ __ __.00

__No Income

__Don't know

7. *INTERVIEWER:*
Record end time of BQ __ __ : __ __

Appendix C:

Model Procedures Manual

TABLE OF CONTENTS

1 INTRODUCTION **5**
- 1.1 INTRODUCTION 5
- 1.2 SURVEY OBJECTIVES 5
- 1.3 OVERVIEW OF DATA COLLECTION METHOD 5
- 1.4 THE SAMPLE 6

2 DIVISION OF RESPONSIBILITY **7**
- 2.1 PROJECT SUPERVISORS 7
- 2.2 SENIOR INTERVIEWERS 7
- 2.3 INTERVIEWERS 8

3 SCHEDULE OF ACTIVITIES **9**
- 3.1 PROPOSED SCHEDULE OF ACTIVITIES 9

4 SUPPLIES **10**
- 4.1 SELF-STUDY SUPPLIES 10
- 4.2 PROJECT SUPERVISOR AND SENIOR INTERVIEWER'S SELF-STUDY 10
- 4.3 SUPPLIES FOR CLASSROOM TRAINING 10
- 4.4 DATA COLLECTION SUPPLIES 11

5 TRAINING **14**
- 5.1 SELF-STUDY TRAINING 14
- 5.2 HEAD OFFICE TRAINING 14
- 5.3 CLASSROOM TRAINING FOR SENIOR INTERVIEWERS AND INTERVIEWERS 14

6 SURVEY PREPARATION **15**
- 6.1 VERIFY SURVEY SUPPLIES 15
- 6.2 MATCH FORMS 02 (CLUSTER LISTS) TO FORMS 01 (CLUSTER DIAGRAMS) 15

6.3 LABELLING 15
6.4 PREPARE INTERVIEWER WORKLOADS 17
6.5 PREPARE SENIOR INTERVIEWER SUPPLIES 18

7 MONITORING DATA COLLECTION 19
7.1 TRACING RESPONDENTS 19
7.2 TRANSFER PROCEDURES 19
7.3 ANSWERING QUERIES 20
7.4 FOLLOW-UP PROCEDURES 21
7.5 PROGRESS AND COST REPORTS 22

8 EDITING 24
8.1 EDITING SURVEY DOCUMENTS 24
8.2 INTERVIEWER EDITS 24
8.3 SENIOR INTERVIEWER EDITS 26
8.4 RE-SCORING OF CORE TASKS 26
8.5 QUESTIONNAIRES RETURNED TO R.O. 27

9 FINAL STATUS CODES 28
9.1 FINAL STATUS CODES 28

10 BATCHING 32

11 SHIPMENT OF SURVEY DOCUMENTS 33
11.1 GENERAL INSTRUCTIONS 33
11.2 SHIPMENTS FROM INTERVIEWERS TO SENIOR INTERVIEWERS 33
11.3 SHIPMENTS FROM SENIOR INTERVIEWER TO THE REGIONAL OFFICE 34
11.4 SHIPMENTS FROM REGIONAL OFFICE TO HEAD OFFICE 34
11.5 MISSING DOCUMENTS 35
11.6 SHIPMENT OF UNUSED DOCUMENTS FROM REGIONAL OFFICE TO HEAD OFFICE 36

12 EVALUATION REPORT 37
12.1 EVALUATION REPORT 37

13 TASK CODES 38
13.1 PROJECT CODE 38
13.2 INTERVIEWER TASK CODES 38
13.3 SENIOR INTERVIEWER TASK CODES 38
13.4 REGIONAL OFFICE TASK CODES 38

1 INTRODUCTION

1.1 INTRODUCTION

The International Adult Literacy Survey is sponsored by the Department of Human Resources Development and the National Literacy Secretariat. This survey will be conducted from September 1st to September 17, 1994.

The survey is a follow-up to the 1989 Survey of Literacy Skills Used in Daily Activities which profiled literacy levels among the Canadian population. The International Adult Literacy Survey will incorporate an international dimension to enable comparison of literacy profiles between countries. Some of the country involved are : United States, Sweden, Switzerland, France, Germany, Netherlands, Poland and Ireland.

1.2 SURVEY OBJECTIVES

Information from this survey will be used by government, researchers and educators to plan program suited to the need of Canadians. Beside the comparison of literacy profiles across international boundaries, the survey will provide an updated profile of adult literacy in both official languages for comparison to that provided by the 1989 Survey of Literacy Skills Used in Daily Activities.

The respondent's information processing skills referring to reading, writing and numeracy necessary to use printed material comonly encountered at work, at home and in the community will be measured.

The survey will also provide a self-assessment of literacy skills and associated literacy needs for persons whose mother tongue is other than English and French. We will obtain a sufficiently large numbers of seniors, recent graduates, and out of school youth to profile their skill levels and allowing the analysis of the link between literacy and other socio-economic indicators.

1.3 OVERVIEW OF DATA COLLECTION METHOD

Interviewers will telephone selected respondents to make an appointment for a personal interview. The interview must be conducted in person in the respondent's home. If the respondent speaks a language other than English or French, the Interviewer should encourage a personal interview with an interpreter present.

The Interviewer will complete the Background Questionnaire with the selected respondent and/or an interpreter (if necessary). The selected respondent will then be asked to complete the Core Tasks Booklet. The data collection tool are paper and pencil. If at least two of six questions in the Core Tasks Booklet are completed correctly (score value = 1), the respondent will be asked to continue the interview and complete one of the seven Main Tasks Booklets. An interpreter cannot help the respondent complete the Core or Main Tasks Booklets.

1.4 THE SAMPLE

The sample of 6,427 respondents were selected. These respondents will be chosen from dwellings in Labour Force Survey rotation groups 2 (rotated-in in February 94) and 3 (rotated-in March 94). Both of these rotation groups will have rotated out of the LFS, one in July and the other in August 1994.

In New Brunswick additional sample was selected from groups 1 (rotated-in January 94), 4 (rotated-in August 94) and rotation group 6 which rotated-in December of 1993. This is being done in order to permit a larger selection of Franco New Brunswickers. Some of the households in these rotation groups may also have been selected for the Youth Smoking survey.

The allocation of the sample by Regional Offices is as follows:

Regional Office	Sample
Halifax	**2,118**
Newfoundland	244 (Base)
Prince Edward Island	129 (Base)
Nova Scotia	415 (Base)
New Brunswick	655 (English sample) 675 (French sample)
Montreal	**1,041** (61 English)
Sturgeon Falls	**485** (35 French)
Toronto	**1,028** (Base)
Edmonton	**1,317**
Manitoba	345 (10 French) Winnipeg R.O. 619
Saskatchewan	444 (Base)
Alberta	528 (Base) Edmonton R.O. 698
Vancouver	**438** (Base)
TOTAL	**6,427**

2 DIVISION OF RESPONSIBILITY

2.1 PROJECT SUPERVISORS

Project Supervisors report directly to the Regional Manager and are responsible for completing those duties delegated by the Regional Manager. Specific duties for this survey include:

- Prepare Interviewer assignments.
- Complete self-study (4 hours).
- Ship self-study packages and assignments to Interviewers and Seniors.
- Mail-out introductory letter to selected respondents.
- Receive classroom training in Head Office (1 day).
- Train Senior Interviewers (1 day).
- Manage data collection:
    - collect progress and cost information;
    - monitor response (conduct follow-up if necessary);
    - monitor costs;
    - ensure all documents have been returned to the R.O.
- Report summary of Progress and Cost to H.O.
- Ensure documents have been batched according to procedures.
- Ship completed documents to H.O.
- Prepare survey evaluation report.

2.2 SENIOR INTERVIEWERS

Senior Interviewers are directly responsible to the Regional Manager/ Project Supervisor for the successful and timely completion of the assignments in their areas of responsibilities. Specific duties for this survey include:

- Complete self-study (4 hours).
- Receive classroom training (1 day).
- Train Interviewers (1 day).

- Supervise data collection:
    - collect progress and cost information;
    - monitor quality of work received from interviewers;
    - monitor response (conduct follow-up by telephone);
    - conduct interviews (if necessary);
    - monitor costs;
    - conduct tracing, if necessary;
    - transfer respondent envelopes to another interviewer or to R.O. if necessary.
- Report summary of progress and cost to Regional Office.

2.3 INTERVIEWERS

Interviewers are directly responsible to the Senior Interviewer to ensure the successful and timely data collection from selected respondents assigned to them. Specific duties for this survey include:

- Complete self-study (4 hours).
- Attend classroom training (1 day).
- Schedule and conduct personal interviews according to procedures in the Interviewer's Manual.
- Score the Core Tasks Booklet completed by the respondent.
- Follow-up non-response.
- Conduct tracing if necessary.
- Edit the Background Questionnaire.
- Assign Final Status Codes to all survey documents.
- Report progress and costs to Senior Interviewer.
- Inform Senior Interviewer of refusals and other non-interviews.
- Inform Senior Interviewer of respondents that have moved outside of the assignment area.
- Review survey progress and discuss any problems or questions.
- Complete administrative forms as required.
- Ship assignment envelopes for which all documents have reached a final status to your supervisor.

3 SCHEDULE OF ACTIVITIES

3.1 PROPOSED SCHEDULE OF ACTIVITIES

ACTIVITY	FROM	TO	WORK DAYS
H.O. provides preliminary counts by assignment number to R.O.s	July 18		
R.O.s select Interviewers for the survey	July 19	July 29	9
H.O. ships self-study materials to R.O.s	Aug. 15	Aug. 17	3
R.O.s ship self-study to Seniors and Interviewers	Aug. 17	Aug. 23	4
R.O. prepares and mails introductory letter to respondents	Aug. 1	Aug. 22	15
H.O. ships remaining materials to R.O.s	Aug. 18	Aug 22	3
Sample file, Assignment lists, Labels files transmitted to Regional Offices	Aug. 8	Aug. 22	10
R.O. prepares Interviewer assignments	Aug. 22	Aug. 26	5
Interviewers and Seniors complete self-study	Aug. 25	Aug. 31	5
Training for Project Supervisors in Head Office (1 day)	Aug. 24		1
Training for Seniors in the R.O. (1 day)	Aug. 29	Aug. 30	1
Training for Interviewers (1 day)	Aug. 30	Aug. 31	1
Data collection period (including Saturdays)	Sept. 1	Sept. 17	14
Clean-up week	Sept. 26	Oct. 1	6

4 SUPPLIES

4.1 SELF-STUDY SUPPLIES

The following is a description of items required for self-study by the Project Supervisors, the Senior Interviewers and the Interviewers prior to classroom training:

Interviewer's Manual – A reference document used by Interviewers for all survey procedures.

Background Questionnaire – The questionnaire to be completed during a personal interview.

Core Tasks Booklet – First respondent booklet containing items which are to be completed by the respondent during a personal interview.

Main Tasks Booklets – Second respondent booklet containing items which are to be completed by the respondent if two or more questions were answered correctly in the Core Tasks Booklet. There are 7 different Main Tasks Booklets. The one to be completed by the respondent will be identified on the label affixed to the Background Questionnaire and on the label affixed to the Respondent's Envelope.

Tasks Administration Guide - A guide containing instructions that are read aloud by the Interviewer to the respondent to complete the Core Tasks Booklet and the Main Tasks Booklet. The Administration Guide is produced in bilingual format and is divided into 8 parts, one for the Core Tasks Booklet and 1 each for the Main Tasks Booklets.

Newspaper – This mock newspaper will be used by respondents to complete certain tasks in the Main Task Booklet.

4.2 PROJECT SUPERVISOR AND SENIOR INTERVIEWER'S SELF-STUDY

Procedures Manual – In addition to the above documents, Senior Interviewers and Project Supervisors must read the contents of this manual.

4.3 SUPPLIES FOR CLASSROOM TRAINING

Instructors will need their Training Guide, their Procedures Manual as well as the self-study package.

Each trainee will require a copy of the Training Workbook for Interviewers. Interviewers should be reminded to bring their Interviewer's Manual with completed self-study exercises and all documents received as part of the self-study package to the class.

4.4 DATA COLLECTION SUPPLIES

The following is a list of items that are needed for data collection:

Procedures Manual – A guide containing survey procedures to be used by Senior Interviewers and R.O. staff to monitor the International Adult Literacy Survey.

Assignment Control List – The list is similar to the Report R22 produced for the LFS. It contains the dwelling identification for respondents that were selected for the survey. It must be used to control the receipt and return of survey documents, and the final status of each.

Head Office will provide three sets: one to be retained by the Regional Office, one for the Senior Interviewer and one to be split among the Interviewers. The list is printed by assignment number and LFS language of interview within each assignment (based on regular LFS assignment). (See annexe 4-A at the of this section).

Respondent's Envelope – One for each selected respondent and in the language indicated on the label. This envelope will contain all the documents required (Background Questionnaire, Core and Main Tasks Booklet, Newspaper) to conduct the interview with the respondent identified on the label of the questionnaire. A final status box will be pre-printed in **black** for the regular survey. The R.O. will affixed a label identical to the label of the Background Questionnaire.

Respondent Labels – Three different labels will be printed in R.O. in the same sequence as the Assignment Control List.

Two of the three labels produced will be identical. One is to be affixed to the space provided on the cover page of the Background Questionnaire, the other, to the upper left-hand corner of the Respondent's envelope. Refer to Section 6.3 for a detailed description of this label.

The third label will be an address label for the mail-out of the introductory letters.

Background Questionnaires – These are to be labelled before they are assigned to Interviewers. Interviewers must receive a pre-labelled Background Questionnaire for each respondent in their assignment. The R.O. should provide Interviewers with spare copies in both official languages.

Core Tasks Booklet – This is the first booklet to be completed by the respondent. One copy in the language indicated on the label will be included in each respondent's envelope.

Main Tasks Booklets – One of the seven different Main Tasks Booklets as identified on the cover of the BQ, must be included in the respondent's envelope. This booklet is to be completed only when the respondent correctly answered 2 of the 6 questions in the Core Tasks Booklet.

Tasks Administration Guides – This guide is divided into eight sections, one for the Core Tasks Booklet and one for each of the Main Tasks Booklets.

Newspaper – The mock newspaper will be used by the respondent to complete some tasks in the Main Tasks Booklet.

Form 01 (Cluster Diagram) – A map provided by the R.O. depicting the area in which a selected respondent is found.

Form 02 (Cluster List) – Forms provided by R.O. for each cluster in which a respondent has been selected.

Final Status Code Sheet – To be used by Interviewers, Senior Interviewers, and Project Supervisors to assign a final status code to the Background questionnaire and the Core and Main Tasks Booklets.

Edit Summary Sheet – To be used by Interviewers, when editing the Background Questionnaire in the respondent's home and after the interview to ensure all required information has been obtained and recorded.

Progress and Cost Reports – There is a Progress and Cost Report for each staff level involved in the survey. Interviewers will receive 1 copy of the Interviewer Progress and Cost Report to record their progress and costs each day. See Section 8 of the Interviewer's Manual for an example of this report.

Senior Interviewers and the Regional Office will receive 4 copies of their reports to summarize the progress and costs of each person they supervise as well as their own progress and costs.

See section 7.4 of this manual for examples of these reports.

Interviewer ID Card – A card designed to show respondents that Interviewers are who they claim to be.

Pencils – To be used by Interviewers and respondents when completing the forms.

Green pens – To be used by Senior Interviewers and Project Supervisors for editing survey documents (No felt pen).

Transmittal Form – A form that accompanies the shipment of all completed questionnaires (see example on page 41).

Literacy Programs List – This list has been compiled by H.O. to assist the Senior Interviewers and Interviewers in providing the names of literacy programs available within the Regional Office boundaries.

ANNEXE 4-A - EXAMPLE OF ASSIGNMENT CONTROL LIST

5 TRAINING

5.1 SELF-STUDY TRAINING

The self-study will be done by Project Supervisors, Senior Interviewers and Interviewers prior to their classroom training.

The self-study materials include:

- the Interviewer's Manual;
- the Background Questionnaire;
- Core Tasks Booklet;
- Main Tasks Booklets (1 to 7);
- Tasks Administration Guide;
- the "newspaper".

The Interviewer self-study should take approximately 4 hours to complete. The review exercises throughout the Interviewer's Manual are to be completed and brought to the classroom training session for review and discussion.

In addition to the Interviewer self-study, Project Supervisors and Senior Interviewers must read the contents of this manual. Allow 2 hours for this task.

5.2 HEAD OFFICE TRAINING

Head Office Training covers Interviewer duties plus survey management responsibilities of the Project Supervisor.

5.3 CLASSROOM TRAINING FOR SENIOR INTERVIEWERS AND INTERVIEWERS

The classroom training for Interviewers will be given to Senior Interviewers by the Project Supervisor. Senior Interviewers will train Interviewers. The Training Guide must be followed to ensure uniform training in all Regional Offices.

6 SURVEY PREPARATION

6.1 VERIFY SURVEY SUPPLIES

As the supplies arrive in the Regional Office, ensure that.

1. You have received the quantities intended for your Regional Office;
2. The quantity shipped is sufficient for your operations;
3. The quality of the material is adequate (for example, spot check questionnaires to ensure they have been printed correctly, ensure packages were not damaged during shipment).

If there are any problems, notify Head Office immediately.

6.2 MATCH FORMS 02 (CLUSTER LISTS) TO FORMS 01 (CLUSTER DIAGRAMS)

The Regional Office will provide the Forms 02 (Cluster Lists) for all clusters in which a respondent has been selected. These clusters are in rotation groups which have rotated out of the LFS sample. The only exception is in New Brunswick which has sample from rotation groups 4.

For every assignment that appears on the R22 "Assignment Control Report", the Regional Office must match the Form 01 (Cluster Diagram) from their files to the corresponding Form 02 so that these may be included in the Interviewers' assignments. It may be necessary to make photocopies of Forms 01 currently in the LFS sample.

6.3 LABELLING

Ensure that two labels like the one below were produced for each dwelling identified on the Assignment Control List. Labels are printed in the same sequence as the Assignment Control List (by language within assignment). The letters "Q" and "E" in the upper right corner of the labels means that the labels must be respectively affixed on the questionnaires and on the envelopes.

The Regional Office must label the Background Questionnaires and Respondent envelopes before the assignment is sent to the interviewer. Affix the label for the respondent envelope in the upper left-hand corner.

An example of the questionnaire and respondent envelope label is shown below:

(1) 17 **(2)** 123456 **(3)** 12345 Q
(4) 12345 12 123 1 123 1 LANG. 1
(5) Joseph M. SEX: 1 AGE 24
Fitzgibbons
(6) Apt. 24,
333 Centertown Road
Literacy, Alta. T2T 3T3
(7) Best time to call

TEL 403 555-1212 **BOOKLET 1**

(1) **R.O. Number** – 2 digit number identifies the Regional Office. For example, "17" would be Edmonton R.O. Ensure that these numbers are transcribed (exactly) onto the Booklets and the Newspaper.

(2) **Docket Number** – 6 digit number. Ensure that these numbers are transcribed accurately onto the Booklets and the Newspaper

(3) **Assignment Number** – 5 digit number representing the assignment number at the time of the Labour Force Survey.

(4) **P.S.U., Group, Cluster, Rotation, Listing and Multiple Number** - the identification numbers used to identify the geographic location of the dwelling. Use these numbers to refer to the Forms 01 and 02 when the listing address on the label is a dwelling description.

LANG - This is the preferred language of the household for the Labour Force Survey. 1 = English, 2 = French, 3 = Either (English or French), 4 = Neither (English or French). The language preference must be confirmed with the selected person when making the appointment.

(5) **Name** – the surname and given name of the selected respondent (i.e., the household member randomly chosen to provide information for the survey). It is possible that a fictitious name was given during the LFS survey. In this case, determine who the respondent is by referring to age and sex which appear just to the right of the given name.

(6) **Address** – the address or description of the dwelling as identified on the Cluster List (Form 02). When making an appointment with the selected respondent, make sure that the respondent is still residing at the listing address identified on the label. Sometimes people move and keep their old telephone numbers. If a respondent has moved, be sure to record the new address. If l'intervieweur lives within 30 miles (50 kilometres) of your home, an appointment may be scheduled.

(7) **Best time to call** – the best time to contact the household for the last month it was in the LFS (Item 44 on the Form 03) will be printed on these two lines.

TEL **Telephone number** – This number will only be provided if it was obtained during an LFS interview. If a telephone number is not available, use the resources available to you to obtain it. If a telephone number cannot be traced, you must visit the dwelling.

BOOKLETS # The number for the Main Tasks Booklet which was selected for this interview will be indicated here.

Labelled questionnaires should be inserted in the corresponding labelled envelope and kept in sequence so that they can be easily separated into Interviewer assignments.

6.4 PREPARE INTERVIEWER WORKLOADS

Use the Assignment Control Lists to separate the sample into Interviewer assignments. An Interviewer's assignment should contain no more than 25 selected respondents.

Identify the name of each Interviewer to which an assignment number has been delegated on each copy of the Assignment Control List. The pre-printed assignment numbers corresponds to the original LFS assignments. When assignments are combined, an interviewer will have more than one assignment number. If assignments are split, photocopy the ACL and indicate the interviewer number change. One copy should be retained in the Regional Office to control the receipt of documents from the Interviewers, one copy should be separated and included in Interviewer's assignment packages and the other copy is for the Senior Interviewer.

Once Interviewer workloads have been established, prepare the contents of the assignment packages. Each Interviewer's assignment package should include:

- Their section of the Assignment Control List.
- A Form 01 and accompanying Form 02 for each cluster identified on the Assignment Control List(s) (when necessary).
- One copy of the Tasks Administration Guide.
- For each respondent whose dwelling is identified on the Assignment Control list, an envelope containing the following in the language indicated on the label:
  - A pre-labelled Background Questionnaire;
  - A copy of the Core Tasks Booklet;
  - The appropriate Main Tasks Booklet as indicated on the questionnaire label;
  - One Newspaper.
- Extra documents for respondents who prefer to be interviewed in the other official language. The regional offices are being supplied with enough extra documents in the other official language.
  - Background Questionnaires;
  - Core Tasks Booklets;
  - Newspapers;
  - Each of the 7 Main Tasks Booklets.

NOTE: **Based on the project supervisor knowledge of the area, the interviewers should be provided with enough extradocuments in the other official language. It is up to provided to each interviewer for their particular assignment.**

- Final Status Code Sheet.
- Pencils for use by the Interviewer and the respondents.
- One Progress and Cost Report.
- Transmittal Forms (about 8 per interviewer).
- One Literacy Programs List (see Section 6.5 below).

6.5 PREPARE SENIOR INTERVIEWER SUPPLIES

Each Senior Interviewer will require the following supplies:

- Copy of the Assignment Control List of each interviewer under their supervision.
- The Senior Interviewer Progress and Cost Report (5 copies).
- Green pens (no felt pen) for editing questionnaires, re-scoring Core Tasks and updating the Transmittal Form.
- Transmittal Forms.
- Spare copies of survey documents in the other official language.
- Extra Core Booklets and Main Task Booklets (English and French).
- Extra newspapers (English and French).

7 MONITORING DATA COLLECTION

7.1 TRACING RESPONDENTS

Tracing will be necessary if you are unable to achieve the response rate objective of 80%. Interviewers should obtain the telephone number (and address, if possible) of any respondent who has moved. If the respondent remains in the Interviewer's assignment area, the Interviewer should arrange to conduct an interview.

Tracing methods are limited for Interviewers. Interviewers should attempt to trace the respondent using the telephone directory, directory assistance and information from anyone living at the selected dwelling. If a respondent has moved but lives within 30 miles (50 kilometres) of an Interviewer's home, then an interview must be attempted.

If tracing determines that the respondent has moved outside of the Interviewer's assignment area, the Senior Interviewer must be contacted immediately so that the interview can be transferred to another Interviewer. (Refer to Section 7.2, Transfer Procedures.)

Because of the short data collection period for this survey, tracing will not be done if the respondent has moved outside the Regional Office area.

NOTE: **The labelled respondent's envelope containing the labelled Background Questionnaire must be transferred to the new Interviewer, Senior or the R.O. Do not create duplicate hand-written documents.**

7.2 TRANSFER PROCEDURES

Interviewer's have been instructed to contact the Senior Interviewer immediately if they have traced a respondent. The Senior Interviewer will decide:

1. If the Interviewer is in the best location to conduct the interview with the respondent. OR
2. If the unit should be transferred to another Interviewer he or she supervises. In this case the interviwer should be instructed to send the respondent's envelope to the Senior Interviewer's home immediately.

Interviewer Action

A. TRANSFER "OUT" PROCEDURES

In this case, the Interviewer who transferred the unit must:

1. Enter all available information (new address, telephone number) regarding the respondent in the "Record of calls" section on the cover page of the BQ.
2. Indicate that this is a "transfer out" in the Record of Calls section on the front cover of the Background Questionnaire and on the respondent's envelope.
3. Draw a line through the identification numbers for this unit on the Assignment Control List and write "transferred out" in the area for comments.
4. Indicate that a unit was "transferred out" of his/her assignment on the Progress and Cost Report (should there be already recorded).
5. Ship the original envelope to the Senior Interviewer with his/her regular shipment of work. **(Item 5 of the BQ will be blank as the unit has not reached a final status.)**

Senior Interviewer Action

A. UNIT TRANSFERRED TO ANOTHER INTERVIEWER UNDER HIS/HER SUPERVISION

The Senior Interviewer who receives the unit will:

1. Transfer the questionnaire to the new Senior Interviewer or another interviewer under her responsibility with instructions to:
    a) Indicate that the unit was "transferred in(to)" the assignment on his/her Assignment Control List and Progress and Cost Report.
    b) Attempt to interview the selected respondent. **(The Interviewer must record the final status codes in Item 5 of the BQ.)**
2. Enter the identification numbers for this unit onto her copy of the Assignment Control List and write "transferred to (name of the interviewer)" in the area for comments.
3. When the new interviewer returns the envelope indicate the final status and that it has been returned from the field on the Assignment Control List.
4. Ship this case to the R.O. with the next shipment.

7.3 ANSWERING QUERIES

Throughout the collection period, questions may be asked by respondents, Interviewers and others.

A. Respondent Queries

Interviewers must be provided with the Regional Office telephone number so that it may be given to respondents who wish to verify the legitimacy of the survey and that the Interviewer is a representative of Statistics Canada. A list of Interviewers should be given to the person(s) in the R.O. who will likely receive these calls.

Respondents may call the Regional Office for information about literacy programs. The Literacy Programs List provides the names of literacy programs available within the Regional Office jurisdiction.

B. Queries from Interviewers/Senior Interviewers

Interviewers may call their Senior Interviewer to resolve problems or answer questions regarding questionnaire content or survey procedures. Interviewers and Senior Interviewers should be encouraged to find the answers to their questions by referring to the Interviewer's Manual or Procedures Manual. If the answers cannot be found and you do not know the answers, call the Project Officer, Linda Bélanger in Ottawa at (613)951-6795. By following the established communication guidelines, other Regional Offices will be informed about potential problems that may arise in their regions.

C. Queries from Others

Interested individuals may request to have copies of the survey documents.

Never distribute copies of the Core Tasks or Main Tasks Booklets. These booklets must remain confidential if we are to accurately profile literacy.

Requests for information from the media should be referred to Jean Pignal the IALS Project Manager in Ottawa at (613) 951-3317.

7.4 FOLLOW-UP PROCEDURES

Senior Interviewers and Project Supervisors must monitor response to ensure that follow-up procedures are implemented to meet the targeted 80% response rate.

Interviewers must report all non-response situations to their Senior Interviewers.

Senior Interviewer Action

Senior Interviewers should advise Interviewers on ways to handle non-response situations. Seniors must also assess each non-interview to determine if follow-up would convert the non-interview status.

Follow-up by Seniors, for the most part, must be done by telephone. Seniors should encourage respondents to participate and, if successful, ask Interviewers to arrange personal interviews. Seniors may conduct personal interviews only if respondents live within a short travel distance from his/her home, or if an Interviewer is not available to conduct the interviews.

Here are some examples of situations that require follow-up:

Refusal: Review the reasons for the refusal. Call the respondent and address his/her specific concerns. Explain why his/her participation is important.

Language problem: If the respondent requests an interview in the other official language, you must attempt to reassign the interview to an Interviewer who speaks that language. If the Interviewer could not arrange an interview because the respondent spoke a language other than English or French, call the household again. Another household member may answer your call and be able to act as interpreter for the background questionnaire. If possible, determine the language spoken and reassign the interview to an Interviewer who speaks that language.

In all cases of language difficulties, the selected respondent should be asked to attempt the Core Tasks Booklet (and the Main Task Booklet if applicable). An interpreter cannot be used for to assist in the completion of either booklet.

No one home: This final status should not be reported until the latter stage of the data collection period, however, you should be aware of this type of non-response throughout the collection period. Ensure that the Interviewers have made calls on various days and at various times throughout the collection period. If an Interviewer has an unusually high number of "no one home" cases, you should call some of the respondents to verify the situation.

Temporarily absent: This status is assigned only if the selected respondent cannot be reached during the entire data collection period.

7.5 PROGRESS AND COST REPORTS

The purpose of Progress and Cost Reports is to provide a system of controlling expenditures and response in relation to the budget and survey progress. These reports will allow Senior Interviewers, Project Managers and Head Office to foresee any potential problems and to take the necessary steps to correct them.

Instructions for completing the Interviewer's Progress and Cost Report are found in Section 8 of the Interviewer's Manual. Instructions for completing the Senior Interviewer and Regional Office Summary Progress and Cost Reports are found on the back of the form.

Interviewer Action

Interviewers must keep a daily record of their progress and cost. Interviewers will report to their Senior Interviewer at the appointed time at the end of each week of the data collection period.

Senior Interviewer Action

Each week the Senior Interviewer must prepare one Senior Interviewer Summary Progress and Cost Report representing the sum total of all of their Interviewers' progress and costs. Any costs incurred by the Senior Interviewer are also to be included on this summary report.

Senior Interviewers will report to the Regional Office once each week. (Refer to the reporting schedule which follows.)

Regional Office Action

Each week, the Project Supervisor must prepare one Regional Office Summary Progress and Cost Report representing the sum total of all of their region's progress and costs.

Regional Offices will report to Head Office the **Tuesday** following each week of data collection. Reports should be sent by facsimile to Linda Bélanger at (613) 951-0582. Refer to the schedule which follows.

Following is the reporting schedule during the survey period:

REPORT NUMBER	FROM INTERVIEWER TO SENIOR	FROM SENIOR TO R.O.	FROM R.O. TO H.O.
1	Tuesday, Sept. 6	Tuesday, Sept. 6	Tuesday, Sept. 6
2	Monday, Sept. 12	Monday, Sept. 12	Tuesday, Sept. 13
3	Monday, Sept. 19	Monday, Sept. 19	Tuesday, Sept. 20
4	Monday, Oct. 3	Monday, Oct. 3	Tuesday, Oct. 4

8 EDITING

8.1 EDITING SURVEY DOCUMENTS

Interviewers are responsible for editing the survey documents before they are shipped to their supervisor.

Head Office will be required to process the data from this survey in within very tight time constraints. When you receive the first batches of questionnaires from Interviewers, ensure that Interviewers are carefully and accurately performing the edits listed boils.

Do not erase entries made by the Interviewer on the Background Questionnaire. Rather, using a green pen, circle the error or omission and speak to the interviewer if necessary. If an item has been left blank, do not attempt to guess the entry that should have been made.

Leave the item blank unless

a) the appropriate entry is obvious, (e.g., interviewer's name) or

b) the appropriate entry can be determined by other information on the questionnaire (e.g., if Item 2, Total number of calls has been left blank, count the number of calls entered in the Record of Calls and Appointments and make the appropriate entry in Item 2.

8.2 INTERVIEWER EDITS

A. EDITS TO BE DONE WHILE RESPONDENT COMPLETES MAIN TASKS

Interviewer Item 4 - Assistance provided

Indicate if the respondent needed assistance for the completion of the Questionnaire.

Interviewer Items 6 and 7 - Start and end times BQ

Ensure that the start and end times for the Background questionnaire were entered in these Interviewer items on pages 3 and 22 of the BQ.

Flows (Go to...)

Ensure that the flows throughout the questionnaire have been correctly followed. Each time you have marked an answer that is followed by a "go to" instruction, circle the instruction to ensure that you have seen it.

Questions B2 to B5

If the answer to B1 is a language other than English (or French on the French questionnaire) there must be entries in questions B2 to B5. If this is the case, ensure that the proper language code(s) have been entered in boxes in questions B2 to B5.

Questions B13 to B17

Ensure that the language(s) reported in B14 to B17 have also been reported in B13.

Other - specify

Review all questions that have an entry in the "Other - specify" category and ensure that the entry you have made does not actually belong in one of the listed categories.

General edits

Ensure that:

- all written entries are complete, correct and legible;
- an "X" has been marked in the appropriate circle;
- the time entries are recorded using the 24-hour clock.

B. FINAL ENTRIES - TO BE DONE AFTER MAIN TASKS COMPLETED

Interviewer Item 8 - Language of Tasks Booklets

Ensure that this item has been marked.

Interviewer Items 9 and 10 - Start and End Time Task Booklets

Ensure that the start and end times for the Tasks Booklets have been entered in these Interviewer items on page ? of the BQ.

Interviewer Item 2 - Total Number of Calls

Verify that the total number of calls from the Record of Calls and Appointments is recorded.

Interviewer Item 3 - Interviewer's Name

Ensure that you have signed your name after editing and recording the final status of the all documents.

Interviewer Item 5 - Final Status

Ensure that a Final Status code has been recorded for all 3 survey documents: the Background Questionnaire and the Core and Main Task Booklets.

Transcribe these codes into the pre-printed Final Status are in the upper right-hand corner of the respondent's envelope.

Non-interview section

If you have not entered Code X or P for any of the documents in Item 5 - Final Status, explain the reason for the non-interview in the Non-Interview section on the last page of the questionnaire.

AFTER THE INTERVIEW, ENSURE THAT ALL THE DOCUMENTS ARE RETURNED TO THE RESPONDENT'S ENVELOPE WHETHER THEY WERE COMPLETED OR NOT.

NOTE: Refer to your Final Status Code Sheet for a summary of the codes for each document. Whenever possible the standard LFS codes have been used.

8.3 SENIOR INTERVIEWER EDITS

The Senior Interviewer's main tasks for the survey will be:

a) to carefully review the first few sets of documents completed by each interviewer and communicate consistent errors to them so that the quality of work can be improved.

b) to randomly review documents received from interviewers throughout the survey period to ensure quality of work is maintained.

c) to ensure that all survey documents (BQ, Core and Main Tasks Booklets and Newspaper), even those that have not been completed are included in each respondent's envelope.

d) to sensure that all the survey documents have the RO Docket Number transcribed onto them.

e) to ensure that each document has been assigned a final status and that this Final Status Code has been transcribed to the pre-printed box in the upper right-hand corner of the respondent's envelope.

f) to re-score 20% of the Core Tasks scored by the Interviewer in Part II of the Background Questionnaire.

8.4 RE-SCORING OF CORE TASKS

The Senior Interviewer will be required to re-score 20% of each interviewer's entries in Part II of the Background Questionnaire, Core Tasks - Scoring. This must be done at the begenning of the survey. Then at various times throughout the survey period for each interviewer. The first 3 questionnaires where the Core Task Booklets was completed should be checked. In this way, any problems that an interviewer is having with scoring will be caught early on. After this, the Senior Interviewer will check every 5th questionnaire until 20% of that interviewer's assignment has been verified. For every 5 questionnaires received from Interviewer's the Senior must remember to re-score 1 of them. Proceed as follows:

- Turn to PART II - Core Tasks - Scoring in the BQ.
- Take out the respondent's Core Tasks Booklet.

- Compare the entry in the Core Tasks Booklet to the possible answers listed beside circles 1, 7 and 0 in Part II of the BQ.
- **In the square box marked S2 in the lower right hand corner of each question block,** enter the code, (1, 7 or 0) that best describes the respondent's entry in the booklet.

NOTE: USE A GREEN PEN TO MAKE YOUR ENTRY IN THE BOX - NEVER ERASE THE ENTRY MADE BY THE INTERVIEWER IN THE CIRCLE.

The Senior Interviewer must discuss with the interviewer if any scoring problems that he/she is having so that they can avoid errors in the core Booklets.

8.5 QUESTIONNAIRES RETURNED TO R.O.

The Regional Office is responsible for the safe return of all documents assigned to Interviewers for data collection.

As shipments are received from Senior Interviewers:

- ensure that the respondent envelopes corresponding to each docket number listed on the Transmittal Form is included in the shipment.
- indicate which documents have been returned on the R.O. copy of the Assignment Control List.

The Regional Office is not responsible for editing questionnaires.

NOTE: You must locate all documents that are not received.

9 FINAL STATUS CODES

Once all documents have been edited, a final status code must be recorded for each type of document: the Background Questionnaire, the Core Tasks Booklet and the Main Tasks Booklet.

When the final status is other than code "X" or "P", review the comments entered in the "Reasons for non-interview" area on the back cover of the Background Questionnaire.

NOTE: Wherever possible, the standard LFS Codes have been used. Refer to your Final Status Code Sheet for a summary of the codes for each document.

The final status codes are defined below. Some codes may seem inappropriate for certain documents. For example you would not normally use Code E, Interview prevented due to blindness or visual impairment on the Background Questionnaire as this questionnaire is administered verbally. However, if the respondent refuses to participate in the survey because he/she is blind, enter Code E on the survey documents for this respondent, not code R (Refused).

9.1 FINAL STATUS CODES

NOTE TO PROJECT OFFICER: REVIEW THESE CODES AND COMPARE TO INTERVIEWERS MANUAL WHEN FINALIZED

Code X: Completed Questionnaire

Background Questionnaire – Contains no more than 5 unanswered questions.

Core and Main Tasks Booklets – There is evidence that the respondent attempted each question. Evidence means that the respondent made a mark on the page . This mark need not be the correct answer. It could be a question mark for example. This is nonetheless evidence that the respondent attempted the question.

Code P: Partial Questionnaire

Background Questionnaire – More than 5 applicable questions were not answered but most of the applicable questions were answered.

If the respondent provides enough information for a "Partial" BQ, attempt to have him/her complete the Tasks Booklets. If the Background

Questionnaire does not meet the requirements for a Partial (i.e., it is a Code R, D, G, H, etc.), end the interview.

Core Tasks Booklets – There is no mark in the booklet for at least 1 or more questions.

If the respondent completed only 2 items on the Core Tasks Booklet and made no entry whatsoever for the other items, the Final Status of the Core Tasks Booklet would be P. If the respondent completed only 2 items and the two answered questions are correct, the respondent should still be asked to complete the Main Tasks.

A partial Core will be scored the same as a completed Core.

Main Tasks Booklet – There is a mark in the booklet next to at least one of the questions.

Code R: Refusal

The respondent refuses to participate in the survey or begins the survey and refuses to continue before the document contains enough information to qualify as a partial. Enter Code R for the document where the refusal occurred and for all subsequent documents.

EXAMPLE: The respondent completed the Background Questionnaire but refused to continue after completing only 2 items in the Core Tasks Booklet. The documents in his envelope would be coded as follows: BQ = X, Core = P, Main = R.

Code T: Temporarily Absent

You have contacted the household but have been informed that the selected respondent will be absent during the entire survey period. Also use Code T when you have reliable information that no household member will be at home during the survey period.

For example:

- a neighbour has told you that all members of the household are on vacation in Florida from September 1st to 30th.
- the selected respondent is away on a business trip and will not be back for three weeks.

In all cases specify the reason and, if possible, the duration of the absence in the "Reason for Non-Interview" section of the questionnaire.

Code N: No one at home

This code applies to the dwelling in which the respondent now lives. Use Code N if, after numerous attempts, you have been unable to contact a member of the household. Also use Code N if you are unable to meet with the respondent. If, however, you are told that the respondent will not be home for the entire survey period, use Code T.

Use Code N also if, on your final visit to a dwelling, you believe that someone is inside but no one answers the door, or you actually see an adult who does not answer the door. DO NOT treat this as a refusal.

Code K: Interview prevented by death, sickness or other unusual circumstances related to the household or to the selected respondent.

Use Code K when unforeseen circumstances, for example a serious but temporary illness that cannot be categorized under disabilities (Codes A to H), death in the household or any other unusual situation prevent the interview from taking place. Other unusual situations include: house in quarantine, household recovering from fire or flood.

Use your best judgement, be tactful and ready to recognize if your presence in the household is detrimental to the interview. A more suitable time might be more productive.

NOTE: **For this survey, language difficulties are not included under code K. See codes 0 and I.**

Code O: Language Difficulties

You were unable to complete the Background Questionnaire because the selected respondent did not speak English or French and an interpreter was not available.

OR

The selected respondent was not able to attempt the tasks in the Booklet because he/she does not have sufficient knowledge of English or French.

Code I: Reading and writing difficulty

Mark this code only when respondents state that they cannot read or write English (or French).

Do not use this code in cases where the respondent reads or writes another language. In those cases, use Code O, Language difficulties.

Code U: Unable to Trace

Use Code U if you are unable to locate the respondent or the household after all tracing attempts have been made.

Code Z: Failed Core Tasks

This code is to be used only on the Main Tasks Booklet if the respondent answered less than two out of the six items correctly on the Core Tasks Booklet. The Core Tasks Booklet would be fully complete (Code X) if the respondent attempted all the items even if they were all incorrect.

NOTE: **If the Core or Main Task Booklets is not administered due to language difficulties, use Code O.**

Code A: Interview prevented due to learning disability

Enter this code if the respondent cannot do the survey due to a learning disability. The most common type of learning disability is dyslexia, a type of impairment that affect a person's ability to read. Another type of learning disability is Attention Deficit Disorder. Persons reporting a learning disability should still be encouraged to attempt the Core and Main Tasks Booklets.

Code B: Interview prevented due to mental or emotional condition

This includes psychological disorders, dementia, autism nervous breakdowns, severe depression, schizophrenia, etc.

Code C: Interview prevented due to mental retardation

The selected respondent is mentally slow. This can be caused by a birth defect, a congenital condition such as Down's Syndrome or accident or injury to the brain that results in brain damage.

Code D: Interview prevented due to hearing impairment

If you encounter a respondent who is hearing impaired you can ask him/her to complete the background questionnaire themselves and to read the instructions for the tasks booklets. Use this code only if the respondent refuses to complete the background questionnaire due to a hearing impairment.

Code E: Interview prevented due to blindness or visual impairment

It will be impossible for respondents who are totally blind to complete the Tasks Booklets. Use this code on those cases. If the respondent refuses to participate in the survey due to blindness enter this code on the background questionnaire as well, not code R.

Code F: Interview prevented due to speech or language impairment

Includes conditions such as severe stuttering and aphasia (impairment of verbal communication caused by brain damage.

Code G: Interview prevented due to physical disability

Physical condition that interfere with the ability to complete the Tasks Booklets include paralysis, amputation of hand/arm, lack of muscle control.

Code H: Interview prevented due to other disability

Includes Alzheimer's disease, senility and any other condition not covered in codes A to G above.

If the wrong booklet number for the Main Tasks Booklet was completed by mistake, do not destroy the completed booklet. Include the booklet with the respondent's other materials.

10 BATCHING

- All documents pertaining to a respondent must be kept together and batched by assignment number.
- Within each batch, envelopes received from Interviewers must be separated as follows:
  - response (respondent envelopes where **at least one** of the survey documents was completed or partially completed) on top.
  - non-response (respondent envelopes where **all** the survey documents are blank, final status T, U, K, R, N or A to H) on the bottom.

Use and elastic band to keep these documents together.

- Record the date that the questionnaire was received from the interviewer on your copy of the Assignment Control List.

11 SHIPMENT OF SURVEY DOCUMENTS

11.1 GENERAL INSTRUCTIONS

All shipments of completed documents are to be accompanied by a Transmittal Form. See example in this section.

IMPORTANT : Ensure the count is always updated.

Completed documents must be securely packaged to comply with the confidentiality requirements of the Statistics Act.

11.2 SHIPMENTS FROM INTERVIEWERS TO SENIOR INTERVIEWERS

Interviewers should be reminded to ship the documents pertaining to the first few interviews they complete to their Senior Interviewer within the first few days of data collection so that the Senior can edit documents as soon as possible and provide feedback to Interviewers who are not following the collection procedures correctly. Thereafter, Interviewers are to ship completed work to their Senior Interviewer twice each week.

The documents should be batched by assignment. The final shipment should include the Assignment Control List(s).

All unused documents are to be returned to the R.O. with the final shipment of work.

After editing and conducting any necessary follow-up the Senior Interviewer will **initial** the Transmittal Form included in the package by the Interviewer before forwarding it with the questionnaires to the Regional Office. **Any changes to the Transmittal Form resulting from editing or follow-up should be made in green pen**.

11.3 SHIPMENTS FROM SENIOR INTERVIEWER TO THE REGIONAL OFFICE

The transmittal form completed by the interviewer must follow the shipment until it gets to Head Office.

The Senior Interviewer will proceed as follows for each shipment of documents received from interviewers:

a) edit the BQ as required using a green pen (no felt pen);

b) re-score Part II, Core Tasks - Scoring for 1 out of every 5 documents received;

c) conduct follow-up as necessary;

d) record the final status code of the documents received on her copy of the ACL;

e) forward the Transmittal Form as completed by the interviewer to the Regional Office; with the survey documents. Any changes required to the Transmittal Form as a result of editing or follow-up must be made in **green pen (no felt pen).**

11.4 SHIPMENTS FROM REGIONAL OFFICE TO HEAD OFFICE

The R.O. will proceed as follows for each shipment of documents received from Senior Interviewers:

a) check that there is an respondent envelope for each docket number listed on each Transmittal Form;

b) record the final status code of the documents received on the R.O. copy of the ACL.

Completed documents must be sent to H.O. as per the following :

1) First shipment of completed questionnaires must arrived in H.O. on September 21, 1994;

2) Second and final shipment (after clean-up) must arrived in H.O. on October 5, 1994.

All completed and non-response documents must be shipped to Head Office. Address boxes to:

Statistics Canada
Operations and Integration Division (Unit C-3)
Section C-12, 2nd Floor
Jean Talon Building
Tunney's Pasture
Ottawa, Ontario
K1A 0T6
ATTENTION: Ron Dubeau

Each box should also indicate:

- the Regional Office name or number
- the number of boxes in a shipment (e.g., 1 of 3, 2 of 3, 3 of 3)

A photocopy of the Transmittal Form as completed by the interviewer is to be sent by Priority Post (separately from the shipment of questionnaires) to Linda Bélanger, ISS, SOD, Jean Talon 6-C6. A second photocopy should be kept by the Regional Office as a Record.

11.5 MISSING DOCUMENTS

The following instructions were designed by the Departmental Security Section. When confidential survey documents are in transit and presumed misdirected or lost these instructions must be followed.

All overdue shipments of confidential documents, whether from the Regional Office or from Interviewers or Seniors, require that tracing action be promptly initiated by the Regional Office, and the facts documented.

Should these confidential documents be unaccounted for on the third business day after tracing action commenced, please take the following action:

- Notify the Chief, Institutions and Social Surveys Section in Head Office, giving the following information.

a) the number of missing documents
b) the interviewer's or Senior's name and address
c) the Carrier involved
d) a summary of tracing action taken to date.

When documenting the facts on misdirected or lost shipments, the Regional Office should have or obtain a copy of the Carrier's Shipping Invoice.

If any misdirected or lost documents for the International Adult Literacy Survey are recovered, please notify Linda Belanger at (613) 951-6795 and ship the documents to:

Statistics Canada
Survey Operations Division
Jean Talon Building, 6-C6,
Tunney's Pasture
Ottawa, Ontario
K1A 0T6

ATTENTION: Linda Bélanger

11.6 SHIPMENT OF UNUSED DOCUMENTS FROM REGIONAL OFFICE TO HEAD OFFICE

All unused Background Questionnaires, Core Tasks Booklets and Main Tasks Booklets, Task Administration Guide and the Newspapers should be returned to Head Office.

Label each box indicating:

- Regional Office name or number
- contents of the box (e.g., unused questionnaires)
- the number of boxes in a shipment (e.g., 1 of 2, 2 of 2).

Send unused documents to Richard Porzuczek. See address on the previous page.

Manuals, Training Guides and Workbooks must be destroyed by shredding or acid wash. **Do not** throw these materials in the garbage.

EXAMPLE OF COMPLETED TRANSMITTAL FORM

12 EVALUATION REPORT

12.1 EVALUATION REPORT

The Regional Manager/Project Supervisor working on the survey will prepare an evaluation report at the end of the data collection period.

As a basic guideline, the report should cover:

- survey preparation;
- training;
- survey procedures;
- quality, progress and cost;
- voluntary nature of survey;
- respondent reactions.

Regional Manager/Project Supervisors may also wish to report on:

- survey materials;
- unusual circumstances;
- problem areas; and
- recommendations.

The Evaluation Report should be forwarded to the attention of Linda Bélanger.

13 TASK CODES

13.1 PROJECT CODE

The project code to be used for the International Adult Literacy Survey is 8176-0.

13.2 INTERVIEWER TASK CODES

03 - Training (self-study and classroom training)

09 - Personal interviewing

10 - Telephone interviewing (includes calling for appointments and tracing)

17 - Assignment preparation

18 - Editing

13.3 SENIOR INTERVIEWER TASK CODES

03 - Training (self-study and classroom training)

08 - Supervision

09 - Personal interviewing

10 - Telephone interviewing (includes calling for appointments and tracing)

17 - Assignment preparation

18 - Editing

13.4 REGIONAL OFFICE TASK CODES

02 - Recruiting Interviewers and Seniors

03 - Classroom Training - receiving

03 - Classroom Training - giving

08 - Supervision

17 - Assignment preparation

18 - Editing

18 - Preparation of documents for Transmittal

Appendix D:

Model Interviewer's Manual

Appendix D:

Model Interviewer's Manual

SELF-STUDY INSTRUCTIONS FOR THE INTERNATIONAL ADULT LITERACY SURVEY

Interviewer training for the International Adult Literacy Survey uses four (4) hours of home-study and one day of classroom training.

Read Sections 1 through 11 of this manual and complete the Review Exercises that are located throughout the manual. These will be reviewed and discussed during the classroom training.

The Interviewer's Manual contains background information, procedures, explanations and definitions that you will require to do your job. Familiarize yourself with your manual and learn how to use it as a reference tool.

Find a quiet place where you can read through this manual without interruption. Keep the following documents by your side to refer to as needed:

- The Background Questionnaire (BQ)
- The Core Tasks Booklet
- Main Tasks Booklets 1 to 7
- The Tasks Administration Guide
- The Newspaper

TABLE OF CONTENTS

1 GENERAL 9
  1.1 Background on the International Adult Literacy Survey 9
  1.2 Survey Objectives 9
  1.3 Overview 10
  1.4 Sample 11
  1.5 Schedule 11
  1.6 Official Language Preference 12
  SECTION 1 - REVIEW EXERCISE 13
2 YOUR JOB AS AN INTERVIEWER 15
  2.1 Job Overview 15
  2.2 Supplies 15
  2.3 Preparation 17
  2.4 Tracing Respondents 18
  2.5 Guidelines for Completing Survey Documents 18
  SECTION 2 - REVIEW EXERCISE 19
3 OBTAINING RESPONDENT COOPERATION 21
  3.1 Initial Contact 21
  3.2 Obtaining Respondent Cooperation 21
  3.3 Survey Operations Division Policy 22
  SECTION 3 - REVIEW EXERCISE 25
4 THE BACKGROUND QUESTIONNAIRE 27
  4.1 Purpose 27
  4.2 Cover Page 27
  4.3 The Background Questionnaire 29
  SECTION 4 - REVIEW EXERCISE 53

5 DESCRIPTION OF THE TASKS BOOKLETS 55
5.1 Introduction 55
5.2 Interruptions during the Interview 55
5.3 Administering the Core Tasks Booklet 56
5.4 Administering the Main Tasks Booklets 57
SECTION 5 - REVIEW EXERCISE 59

6 EDIT STEPS 61
6.1 Edits to the Background Questionnaire 61
SECTION 6 - REVIEW EXERCISE 63

7 FINAL STATUS 65
7.1 General 65
7.2 Final Status Codes 65
SECTION 7 - REVIEW EXERCISE 69

8 PROGRESS AND COST REPORT 71
8.1 Progress and Cost Report - Interviewers 71
8.2 Interviewer Worksheet 72
8.3 Example of Completed Progress and Cost Report 73

9 TASK CODES 75
9.1 Task Codes 75

10 TRANSMITTAL OF DOCUMENTS 77
10.1 Transmittal of Documents 77
10.2 Example of Completed Transmittal Form 78
10.3 Transfer procedures 79

11 QUESTIONS AND ANSWERS 81
11.1 Introduction 81
11.2 How did you select me for this survey? 81
11.3 Why me instead of someone else in the family? 81
11.4 Am I obligated to take part in the survey? 81
11.5 Can we do the interview over the telephone? 82
11.6 I'm too old. Why don't you talk to my son; he could better answer your questions? 82
11.7 How much time will the interview take? 82
11.8 Who is involved in conducting this survey? 82
11.9 How many people are included in the survey? 82

11.10 Do you need this information? Couldn't you get it from another source? 82
11.11 The Government has too much information already. 82
11.12 Who uses the information? 82
11.13 What is the International Adult Literacy Survey? 83
11.14 These surveys are a waste of taxpayer's money. 83
11.15 How will I know if all this will be kept confidential? 83
11.16 Does Statistics Canada release information about individuals to other agencies? 83
11.17 Can I get a copy of my Task Booklets? 83
11.18 When will the results be published? 83
11.19 How will the results be given to the general public? 83
11.20 I thought that a literacy assessment had already been done? Why are you doing another one? 84
11.21 What do you mean by literacy? 84
11.22 Why are people paying so much attention to adult literacy? 84
11.23 Are most adults with literacy problems immigrants? 84
11.24 How can Canadians have literacy problems when we have compulsory schooling in Canada? 84
11.25 Didn't the federal government recently give a lot of money to literacy programs? What has happened with that? 84
11.26 What programs are available to improve my skills (Who can I talk to about literacy programs)? 84

APPENDICES 85

APPENDIX A: FINAL STATUS CODE SHEET 87
APPENDIX B: EDIT SUMMARY SHEET 88

1 GENERAL

1.1 Background on the International Adult Literacy Survey

There is a growing concern about the present and future economic and social costs related to the literacy abilities of the Canadian population. In response to this concern, Statistics Canada was commissioned by the National Literacy Secretariat (NLS) and Human Resources and Labour to undertake the Survey of Literacy Skills Used in Daily Activities, a national literacy survey which was conducted in 1989.

As a follow up, Statistics Canada has been asked to participate in the **International Adult Literacy Survey**.

As the title suggests the survey will incorporate an international dimension to enable comparison of literacy profiles between countries. Some of the countries involved are: the United States, Sweden, Switzerland, France, Germany, the Netherlands, Poland and Ireland.

A pilot test of this survey was conducted in October of 1993. Many recommendations from the pilot survey have been implemented for the main survey.

1.2 Survey Objectives

The objectives of this survey are as follows:

- to provide data which can be compared across international boundaries;
- to provide an updated profile of adult literacy in both official languages for comparison to that provided by the 1989 Survey of Literacy Skills Used in Daily Activities;
- to provide a self-assessment of literacy skills and associated literacy needs for persons whose mother tongue is other than English or French;
- to provide sufficiently large numbers of seniors, recent graduates, and out of school youth to profile their skill levels; and,
- to provide data allowing the analysis of the link between literacy and other socio-economic indicators such as educational attainment, labour market participation and employment.

For the purpose of this survey, literacy has been interpreted as "official language literacy" and defined as:

> "The information processing skills (referring to reading, writing and numeracy skills) necessary to use printed material commonly encountered at work, at home and in the community."

Literacy, in this survey's context, will refer to literacy in either one of Canada's official languages.

Literacy proficiency will be measured using three literacy scales:

Prose Literacy - the knowledge and skills needed to understand and use information from texts such as editorials and news stories.

Document Literacy - the knowledge and skills required to locate and use information contained on schedules, maps, tables, graphs, indexes, and so forth.

Quantitative Literacy - the knowledge and skills needed to apply arithmetic operations, either alone or sequentially, that are embedded in printed materials, such as calculations required on deposit slips, amount of time from schedules, cost of goods.

Information from this survey will be used by government, researchers and educators to plan programs suited to the needs of Canadians.

1.3 Overview

Adults with low literacy abilities often feel powerless because they must depend on others to help them with many of the duties we assume as adults. These include such activities as banking, filling in forms, reading dosage information on medicine bottles and signing a lease for a house or apartment.

Because society tends to place a high value on literacy, some adults with literacy problems find it uncomfortable to identify themselves and develop a number of defence strategies to prevent discovery. It will be your task to convince these individuals to take part in the survey.

Some respondents may find the survey threatening. Unlike other surveys, the respondent may perceive that there are "right" and "wrong" answers. This may be the respondent's first exposure to a written test in many years.

It is important that you present yourself as non-threatening and non-judgemental, and that the respondent realizes that he/she is in control of the situation.

Your first contact with the respondent will be on the telephone. Interviewers will telephone selected respondents to introduce the survey and to make an appointment for a personal interview. The interview must be conducted in person, in the respondent's home, rather than on the telephone.

In the respondent's home you will first administer the Background Questionnaire. If the respondent is unable to speak English/French an interpreter can be used to translate the questions from the Questionnaire for the respondent, and the answers for the Interviewer. This situation must be indicated by marking "Yes" in Item 4 on the cover page of the Background Questionnaire.

Once the Background Questionnaire has been administered the respondent is then asked to complete the Core Tasks Booklet. This booklet contains 6 items which assess the respondent's basic literacy level. When the respondent has completed the Core Tasks Booklet, you will score each item.

If the respondent successfully completes 2 of the 6 items, he/she will be asked to complete one of the Main Tasks Booklets, as indicated on the label of the Background Questionnaire. If the respondent does not successfully complete 2 items in the Core Tasks Booklet, the interview ends.

1.4 Sample

The sample for the survey is approximately 6,500 selected respondents. These respondents will be chosen from dwellings in Labour Force Survey rotation groups 2 and 3. Both of these rotation groups will have rotated out of the LFS sample, one in July and the other in August.

In New Brunswick an additional sample will be selected from three additional rotation groups (groups 6, 1 and 4). This is being done in order to permit the selection of a Franco New-Brunswick sample.

NOTE: Some of the households in rotation groups 1, 4, and 6 for New Brunswick may have also been selected for the Youth Smoking Survey to be held in October. Please do not tell respondents that this will be their last contact with Statistics Canada.

The sample will be representative of the Canadian population. It will be stratified by province (urban and rural representation). Following is the distribution by Regional Office:

REGIONAL OFFICE	SAMPLE
Atlantic	2,118
Quebec	1,041
Ontario	1,531
Prairies	1,317
Pacific	438
TOTAL	6,427

1.5 Schedule

ACTIVITIES	DATE FROM	DATE TO
Training - Self-Study, review exercises (4 hours)	August 25, 1994	August 31, 1994
- Classroom Training (1 day)	August 30, 1994	Sept. 1, 1994
- Data Collection (14 days, including Saturdays)	September 1, 1994	September 17, 1994
- Clean-up week (6 days)	September 26, 1994	October 1, 1994

1.6 Official Language Preference

Due to the nature of this survey, it is imperative that you determine the respondent's language preference for conducting the interview. If, part way through the interview for the Background Questionnaire the respondent indicates that he/she would be more at ease in the other official language you must continue the interview in the other language if you can and transcribe the answers to the previous questions onto the new questionnaire after the interview.

If however, the respondent wishes to change the language of the test after you have begun administering the Tasks Booklets, you must re-administer the booklet(s) in the respondent's preferred official language. The same Booklet numbers, as indicated on the label of the Background Questionnaire, must be used whether in English or French.

SECTION 1 - REVIEW EXERCISE

Circle the letter beside the correct answer.

1. The main objectives of the International Adult Literacy Survey are:

   (a) to provide data which can be compared across international boundaries;

   (b) to provide an updated profile of adult literacy in both official languages for comparison to that provided by the 1989 Survey of Literacy Skills Used in Daily Activities;

   (c) to provide a self-assessment of literacy skills and associated literacy needs for persons whose mother tongue is other than English or French;

   (d) to provide sufficiently large numbers of seniors, recent graduates, and out of school youth to profile their skill levels; and,

   (e) to provide data allowing the analysis of the link between literacy and other socio-economic indicators such as educational attainment, labour market participation and employment;

   (f) All of the above.

2. If the respondent attempts the six questions in the Core Tasks Booklet and gets only two correct items, he/she will be asked to complete the Main Tasks Booklet.

   (a) True

   (b) False

3. A respondent does not speak English or French. They are asked to:

   (a) Answer all the Background Questionnaire without assistance

   (b) Have someone act as interpreter to assist in completing the Background Questionnaire

   (c) Have the interpreter assist in completing the Background Questionnaire and the Booklets

4. You will be surveying a sample of dwellings that are part of the:
    - (a) Southam Literacy Survey
    - (b) Labour Force Survey
    - (c) General Social Survey

5. The Background Questionnaire and the Core and Main Tasks Booklets must all be completed in the same official language.
    - (a) True
    - (b) False

6. The Core Tasks Booklet and the Main Tasks Booklet must be completed in the same official language.
    - (a) True
    - (b) False

2 YOUR JOB AS AN INTERVIEWER

2.1 Job Overview

Your job as an Interviewer is not confined strictly to interviewing respondents. While this aspect of your job is very important, your job is diversified and includes the following:

- Completing a self-study program and attending classroom training
- Interviewing selected respondents
- Scoring the Core Tasks Booklets and administering the Main Tasks Booklets
- Completing Progress and Cost Reports

More specifically your job consists of:

1. Telephoning each selected respondent to schedule a personal interview.
2. Completing a Background Questionnaire with the selected respondent during a personal visit.
3. Using the Task Administration Guide, you will read instructions to the selected respondent so that he/she can complete the Tasks Booklets.
4. Scoring the Core Tasks Booklet using Part II of the Background Questionnaire.
5. Editing each Background Questionnaire to ensure that it has been completed properly.
6. Maintaining a daily Progress and Cost Report.
7. Reporting progress and cost to your supervisor.
8. Shipping the completed documents to your supervisor.

2.2 Supplies

Your assignment will include the following supplies:

- **Assignment Control List(s)** - This(these) list(s) will identify your workload.

- **Cluster Diagrams (Form 01) and Cluster Lists (Form 02)** for each cluster in which a respondent was selected.
- **Introductory letter** - extra copies of this letter which was mailed to the respondent by the Regional Office.
- **Respondent's envelope** - For each selected respondent you will receive an envelope containing a labelled Background Questionnaire, a Core and a Main Tasks Booklet and a Newspaper. A label identical to the one on the Background Questionnaire will be affixed in the upper left-hand corner of the envelope. A Final Status Code box will be pre-printed in the upper right-hand corner. You must transcribe the Final Status Code of each document from Item 5 on the Background Questionnaire into this box.
- **Background Questionnaire (BQ)** - One for each respondent identified on the Assignment Control List. Questionnaires will be labelled according to the preferred language of the household for the Labour Force Survey.

  The Background Questionnaires will be administered by the Interviewer during a personal visit to the selected respondent. If the respondent cannot speak either official language, the Questionnaire may be completed with the assistance of an interpreter. Alternatively, if the Interviewer speaks the language of the respondent, he or she may act as interpreter for completion of the Background Questionnaire.
- **Core Tasks and Main Tasks Booklets** - These tasks will be attempted by the selected respondent after the Background Questionnaire has been administered. Respondents with language problems should be encouraged to attempt to complete the Tasks Booklets. No interpreter will be allowed to assist in the completion of the tasks included in the Tasks Booklets.

  The six screening questions, contained in the Core Tasks Booklet, are asked of all respondents. The Core Tasks will be scored immediately after the respondent has finished them. If two (2) or more of these six questions are answered correctly, the respondent is asked to complete the Main Tasks Booklet.

  Each respondent's envelope will include 1 of the 7 Main Tasks Booklet randomly selected for that respondent as indicated on the questionnaire and envelope labels. It is very important to ensure that the respondent completes only the Main Tasks Booklet indicated on these labels.
- **Newspaper** - The newspaper will be used by the respondent to accomplish certain tasks in the Main Tasks Booklet. The questions involving the newspaper are answered either in the Main Tasks Booklet or directly on the newspaper. One newspaper is used for each selected respondent.
- **One Task Administration Guide** - The Task Administration Guide will provide you with the instructions to follow in order to administer to the respondent the Core Tasks Booklet and the Selected Main Tasks Booklet. This guide will be divided into 8 clearly identified sections, one for each booklet.

- Transmittal Forms.
- One Progress and Cost Reports.
- One Final Status Code Sheet.
- One Edit Summary Sheet.
- Blank documents for respondents who wish to be interviewed in the other official language.
- List of Provincial/Municipal Literacy offices when available. This list is a reference to be used if the respondent requests information on literacy programs.
- Pencils (may be left with respondents).

2.3 Preparation

When you receive your assignment:

a) Ensure that each respondent's envelope contains all the required documents;

b) Ensure that the Main Tasks Booklet in the respondent's envelope is the one identified on the envelope and Background Questionnaire labels;

c) **Transcribe the R.O. and the docket number from the label of the Background Questionnaire onto all documents in the envelope.** Use a pencil so that documents can be re-used if necessary.

Review your assignment by noting the location of each respondent in relation to your own home.

Plan personal visits to make efficient use of travel time. It may be best to visit dwellings closest to your home first and travel outwards from there. If there was no response on the first attempt it will be more efficient to try again on your return trip home.

You should plan to spend the day portion of the first day, and the evening portion of the second day conducting telephone calls. By devoting this time early in the collection period, you should be able to schedule appointments for the personal visits. Your initial calls will also determine if a respondent has moved.

When calling to make appointments, state your name and the name of the organization. Be prepared to provide an introduction to the survey in case the respondent has not received the introductory letter mailed by the Regional Office. Try to establish a time suitable for both you and the respondent. Avoid making appointments at specific times. Rather, try to make appointments that allow for unforeseen circumstances (for example, an interview took longer than expected, traffic was congested due to an accident or bad road conditions). Ask the respondent, for example, if he will be home during the morning, rather than making an appointment for 10:00 a.m.

If a respondent requests to be interviewed in the official language you do not speak, inform your supervisor immediately so that appropriate action can be taken.

Plan to complete as many interviews as possible during the first week of the data collection period. Remember that respondents are more easily found at home early in the week.

Include the Task Administration Guide and the list of literacy offices with your interviewing kit. In bilingual assignment areas ensure that you have blank forms in the other official language. Being organized will help the interview flow smoothly and you will project a professional image.

2.4 Tracing Respondents

As the sample was chosen from dwellings that have just rotated out of the Labour Force Survey sample, it is unlikely that respondents would have moved and will need to be traced.

Tracing methods are limited for Interviewers. You should attempt to trace the respondent using the telephone directory, directory assistance and information from anyone living at the selected dwelling. If a respondent has moved but lives within 30 miles (50 kilometres) of your home, then an interview must be attempted. If you are successful in tracing a respondent but he or she lives outside the 30 mile (50 km) limit, contact your Senior Interviewer immediately. She will re-assign the case to another interviewer.

All steps taken to trace the respondent must be entered in the "Record of Calls/Visits" section of the Background Questionnaire.

2.5 Guidelines for Completing Survey Documents

When completing the Background Questionnaire follow these guidelines:

- Use pencil only.
- Record times using the 24 hour clock (e.g. 8:30 a.m. = 08:30 hrs; 4:30 p.m. = 16:30 hrs).
- Print clearly and neatly in BLOCK LETTERS. Be especially careful when making similar shaped letters such as D and O; C and G; U and V.
- Answer questions by placing an "X" in a circle or by clearly printing the respondent's answer in the space provided.
- Erase completely any mistakes you make during the interview so that your entries will not be misinterpreted.
- Use the "Comments" area of the questionnaire to record unusual situations or to provide any further information you feel might be important.
- If the respondent does not know or refuses to answer a question, write "DK" or "R" under the item number and explain the reason for the refusal in the Non-interview section on the back cover of the questionnaire.
- When you transcribe the R.O. Number (2 digits) and the Docket Number (6 digits) from the label on the Background Questionnaire, onto the Core Tasks Booklet, the Main Tasks Booklet and the Newspaper, ensure that you transcribe the exact numbers.
- You will read aloud to the respondent all the questions and categories printed in bold letters. Instructions in italics are for your information. Text in parentheses is supplementary: read it only if you feel further explanation is required.

Following these guidelines will ensure that the questionnaire has been completed properly.

SECTION 2 - REVIEW EXERCISE

Circle the letter beside the correct answer or enter the correct response.

1. When you receive your assignment, transcribe the R.O. and Docket number from the label on the Background Questionnaire to all the other supplies in the respondent's envelope. Each respondent's envelope will contain a Background Questionnaire and

   (a) ______________________________

   (b) ______________________________

   (c) ______________________________

2. Indicate at least three tracing methods/tools you would use to locate a respondent who has moved?

   (a) ______________________________

   (b) ______________________________

   (c) ______________________________

3. How would the following times be recorded in the Record of Calls on the Background Questionnaire?

   (a) 10:30 in the morning ____________________

   (b) 12:00 noon ____________________

   (c) 4:15 in the afternoon ____________________

   (d) twenty to eight in the evening ____________________

4. Which of the following should you read aloud to the respondent?

   (a) Phrases in italics

   (b) Interviewer Check Items

   (c) Questions and categories printed in bold letters

5. When do you mark "R" under a question number?
   (a) If a respondent refuses to be interviewed during initial contact
   (b) When a respondent refuses to answer a question on the BQ
   (c) If a respondent cannot be contacted

3 OBTAINING RESPONDENT COOPERATION

3.1 Initial Contact

The initial contact will be by telephone: Introducing yourself and convincing the respondent to grant an interview are probably the most important and difficult parts of interviewing.

It is extremely important that your voice and words convey your credibility. You should be polite, pleasant, self-confident and at all times professional. If you sound uncertain this will be detected by the respondent who may doubt the validity of the survey. This can be overcome by assuming the respondent is friendly and interested in the survey and its outcome. If the respondent does not appear to be receptive it may be that he/she needs more information about the survey.

Follow these guidelines for a good introduction:

(a) Begin by identifying yourself and stating that you work for Statistics Canada.

(b) Some respondents will grant you the information with only a brief explanation, others will need more detail. Use the Questions and Answers in Section 11, of this manual.

(c) The introduction should not contain any questions that might invite an undesired response. For example, don't ask for permission to do the survey but make a statement that assumes the respondent is willing. For example, "Would it be convenient to meet on .... at ....?"

(d) Do not hesitate between questions or statements. A pause may invite a refusal.

When scheduling an appointment, ensure that the time is convenient for you and the respondent. Explain that the interview could last about one hour.

3.2 Obtaining Respondent Cooperation

Survey information, as identified on the Background Questionnaire, and the Core and Main Task Booklets should be collected from each selected respondent. However, due to refusals and situations whereby you are unable to contact the respondent, this is not always possible.

There is no sure method of eliminating non-response, but the ability to gain cooperation from all types of persons, and planning calls when people are most likely to be home are important factors.

During the personal visit, ensure that the respondent has time to voice his/her opinions and to ask questions. Your answers should be prompt and accurate. For this reason it is imperative that you are familiar with all the documents and with Section 11, Questions and Answers.

3.3 Survey Operations Division Policy

Survey Operations Division policy concerning contact with the public must be followed. The main points of the policy are summarized below.

(a) The International Adult Literacy Survey is a voluntary survey and is dependent upon the respondent's cooperation to participate in the survey. Respondents are under no legal obligation to participate.

Although Statistics Canada has the authority to collect information, it is our policy to rely on the cooperation of the public to obtain it. Our objective is the collection of reliable information and, for this, we need the cooperation of the public.

(b) Information collected is kept strictly confidential. Information is never released or published in any way that could identify an individual. Statistics Canada personnel, including Interviewers, must take an oath of confidentiality and are subject to prosecution if they divulge information related to any identifiable respondent.

Follow these guidelines to gain the respondent's cooperation during the initial contact and personal interview:

(a) Explain the purpose of the survey and uses of the data collected.

(b) Emphasize the importance of the respondent's participation in the survey.

(c) Assure the respondent of the confidentiality of the information provided.

(d) Answer any questions the respondent may ask. (Refer to Section 11 for answers).

(e) Do not consider the reply "I am too busy at the moment" as a refusal. Instead, offer to call back at a more convenient time.

(f) Set your appointments for personal visits, giving yourself adequate time to keep all other appointments. Use the Cluster Diagrams and the Cluster Lists to plan interviews. Before leaving for a scheduled appointment, call the respondent to verify that he/she will be at home. An Interviewer Worksheet is provided on the back page of the Interviewer's Manual to be used for logging in hours worked and distance travelled.

(g) Use the techniques you learned from the booklet, "Introduction to Interviewing."

If you contact a selected respondent who objects to providing the information even after you have explained the purpose of the survey and have made an effort to gain the respondent's cooperation, note the reason for the refusal in the "Reason for Non-Interview" section of the Background Questionnaire. Notify your supervisor daily of all refusals and other non-response.

Be sensitive to respondents who are facing personal or family hardships. Use your judgement to determine if the interview might be better done at another time, or not at all.

SECTION 3 - REVIEW EXERCISE

Circle the letter beside the correct answer.

1. The International Adult Literacy Survey i[illegible] voluntary survey.
   (a) True
   (b) False

2. The respondent tells you he is busy at the moment. What would you do?
   (a) Code the Background Questionnaire as a refusal
   (b) Make an appointment to call at a more convenient time
   (c) Offer to conduct the interview by telephone

3. You will notify your supervisor daily of all refusals.
   (a) True
   (b) False

4 THE BACKGROUND QUESTIONNAIRE

4.1 Purpose

The literacy skills of individuals are to a large extent influenced by early background environment. To gain deeper insights into the literacy abilities of Canadians, information is collected on personal and parental education, labour force, use of mother tongue in reading and writing, ethnic origin, personal and household incomes.

4.2 Cover Page

LABEL

A label containing the following information will be affixed to the cover page of the questionnaire and to the respondent's envelope.

(1) 17 **(2)** 123456 **(3)** 12345
(4) 12345 12 123 1 123 1 LANG. 1
(5) Joseph M. SEX: 1 AGE 24
Fitzgibbons
(6) Apt. 24,
333 Centertown Road
Literacy, Alta. T2T 3T3
(7) Best time to call
TEL 403 555-1212 **BOOKLET 1**

(1) **R.O. Number** - 2 digit number identifies the Regional Office. For example, "17" would be Edmonton R.O. Ensure that these numbers are transcribed (exactly) onto the Booklets and the Newspaper.

(2) **Docket Number** - 6 digit number. Ensure that these numbers are transcribed accurately onto the Booklets and the Newspaper

(3) **Assignment Number** – 5 digit number representing the assignment number at the time of the Labour Force Survey.

(4) **P.S.U., Group, Cluster, Rotation, Listing and Multiple Number** – the identification numbers used to identify the geographic location of the dwelling. Use these numbers to refer to the Forms 01 and 02 when the listing address on the label is a dwelling description.

LANG - This is the preferred language of the household for the Labour Force Survey. 1 = English, 2 = French. You must confirm the language preference of the selected person when making the appointment.

(5) **Name** – the surname and given name of the selected respondent (i.e., the household member randomly chosen to provide information for the survey). It is possible that a fictitious name was given during the LFS survey. In this case determine who the respondent is by referring to age and sex which appear just to the right of the given name.

(6) **Listing Address** – the address or description of the dwelling as identified on the Cluster List (Form 02). When making an appointment with the selected respondent, make sure that the respondent is still residing at the listing address identified on the label. Sometimes people move and keep their old telephone numbers. If a respondent has moved, be sure to record the new address. If the respondent lives within 30 miles (50 kilometres) of your home, you may schedule an appointment.

(7) **Best time to call** – the best time to contact the household for the last month it was in the LFS (Item 44 on the Form 03) will be printed on these two lines.

TEL Telephone number – This number will only be provided if it was obtained during an LFS interview. If a telephone number is not available, use the resources available to you to obtain it. If a telephone number cannot be traced, you must visit the dwelling.

BOOKLET # The number of the Main Tasks Booklet which was selected for this interview will be indicated here.

2. TOTAL NUMBER OF CALLS

Record the total number of calls, including the personal visits/interviews.

3. INTERVIEWER NAME

The name of the Interviewer must be:

a) the Interviewer who finished the interview with the selected respondent; or

b) the Interviewer who obtained the non-response.

4. ASSISTANCE PROVIDED

We want to know if the respondent needed help from a third party to complete the background questionnaire. Check "yes" or "no".

5. FINAL STATUS

Enter the final status of the Background Questionnaire, the Core Tasks Booklet and the Main Tasks Booklet in this space.

NOTE: **A final status code must be entered for all documents in the respondent's envelope in all cases. For example, if the respondent refuses the BQ, you must enter R on the BQ, the Core Tasks Booklet and the Main Tasks Booklet.**

RECORD OF CALLS AND APPOINTMENTS

This section will provide you with a record of all calls made in order to reach the respondent and the outcome of these calls. If unable to contact the respondent, specific tracing steps must be taken and recorded here. (See Tracing Respondents - Section 2.4). Also indicate any rescheduling of interviews.

DATE	enter the date of each call
START TIME	enter the time you called the respondent, used tracing methods or started the personal interview
FINISH TIME	enter the time the telephone call, tracing method or personal interview ended
NOTE	enter all information pertinent to the call, such as:

- appointment date and time scheduled with respondent
- selected person not in - call back (indicate date and time)
- no answer, busy, answering machine, no service at that number
- household moved
- wrong number
- selected person deceased
- unable to reach

* When speaking with the household respondent verify the language in which the interview will be conducted - English or French. If it is determined that the selected person does not speak English or French, ask the respondent if an interpreter is available to assist in the completion of the Background Questionnaire.

INTRODUCTION

You must read aloud the introduction to ensure that the respondent understands the intent and the confidentiality of the survey.

4.3 The Background Questionnaire

6. Interviewer – Record start time of interview

Be sure to complete this Interviewer Item on the second page of the questionnaire before you ask question A1. Use the 24 hour clock when recording the start time of the interview.

SECTION A. GENERAL INFORMATION

QUESTION A1 – First I'd like to ask you a few questions about your background, your education, the languages you speak and jobs you may have held in the past 12 months.

Were you born in Canada?

Mark "YES" if the respondent was born in Canada or Newfoundland before it joined Confederation.

Mark "NO" if the respondent was born outside Canada to Canadian parents who registered the child as Canadian at the time of birth.

QUESTION A3 – In what year did you first immigrate to Canada?

If a respondent's parents were Canadian citizens but he/she was born in another country the respondent may have Canadian citizenship (Foreign Service or Armed Forces personnel.) If the respondent indicates this, mark the circle "Canadian citizen by birth" and go to Question A7.

QUESTION A4 – In total how many years have you lived in Canada?

Include the total number of years. If the respondent lived in Canada for 10 years then left for 5 and has been back for 10, enter 20 in this space.

QUESTION A5 – Before you first immigrated to Canada, what was the highest level of schooling you had completed?

Make sure the respondent has reported the highest grade completed, not attended.

Accept only one response. If the respondent finds it difficult to equate his education to the Canadian system, ask him/her to provide an estimate of the equivalent grade.

If the respondent has received a degree (Bachelor's, Master or PhD), you would mark "Completed university". Always mark the level that the respondent perceives as the highest completed.

A respondent may indicate that he/she took courses in Word Processing and Office Management at the local college. In this case you would probe to determine if he/she completed the requirements for a college certificate or diploma and mark "Completed non-university Post-Secondary".

As well, if a respondent has completed CEGEP (in Quebec only), you would mark "Completed non-university Post-Secondary".

No education: Mark this category if the respondent never attended school or received no education of any type.

Did not complete primary: Mark this if the respondent did not complete grade 6.

Completed primary: refers to the first six to eight grades in the educational system. If a respondent indicates that he/she completed grade 7 or 8, determine if this was completed at elementary or secondary (Junior high) school.

Some secondary: Mark this category if the respondent attended high school but did not graduate from the program in which he/she was registered.

BEST COPY AVAILABLE

If the respondent says that he/she completed high school, determine the type of program and mark the appropriate category below, "Completed secondary (vocational or technical) and "Completed secondary (General and academic)".

Completed secondary (vocational or technical) – Technical/Vocational Programs refers to those schools which offer courses specific to a trade or vocation (such as electronics, automobile mechanics, etc.). These programs can lead to college but not to university.

Completed secondary (general or academic) – A general/academic type of program usually includes courses such as languages, mathematics and the sciences at a general or advanced level to prepare the student for further education at a college or university. These programs **do not** prepare the student for a trade or vocation.

Completed non-university post-secondary – This category includes:

- Community Colleges, etc. – refers to an institution of higher learning offering instruction in a professional, vocational or technical field. A college generally grants diplomas or certificates but not degrees.
- Teacher's colleges
- Technical Institute
- Schools of Nursing
- "CEGEP" – "Collège d'enseignement général et professionnel" is a community college in the province of Québec.

Completed university – A university is an institution of higher learning providing facilities for teaching and research and authorized to grant academic degrees. A university is usually made up of an undergraduate division which confers bachelor's degrees, a graduate division which comprises a graduate school and professional schools each of which may confer masters degrees and doctorates. Mark this category if the respondent received an undergraduate degree.

Completed university post-graduate – Mark this category if the respondent received a degree above the undergraduate level, for example a Masters or PhD.

Education not definable by level – Mark this category for situations such as the following. The respondent:

- attended school in another country and is unable to convert to a Canadian equivalent OR
- attended an ungraded school OR
- studied under a tutor.

QUESTION A8 – What is the highest level of schooling you have ever completed?

Refer to Question A5 for explanation of schooling categories.

QUESTION A9 – Which of the following best describes your secondary program? Was it ...

Read items aloud to the respondent.

An academic/college preparatory type of program usually includes courses that will prepare a student for further education at a college or university. The program will often include courses in mathematics and the sciences at a general or advanced level.

A business (commercial) or trade/vocational type program usually includes courses that will prepare a student for jobs in the business or trade/vocational market (for example – secretarial work, woodworking or machine shop operator). Courses in this program may be directed at a basic level of learning rather than a general or advanced level.

A second level, second stage equivalency program – Mark this if the respondent has completed a High school equivalency program.

QUESTION A10 – In what province or country did you complete your secondary education?

The province or country reported should be based on current boundaries.

QUESTION A11 – Which of the following best describes your secondary program? Was it ...

Refer to Question A9.

QUESTION A12 – What was the main reason you stopped your schooling when you did?

Do NOT read the list of reasons to the respondent, but assist if they have difficulty. Mark the item that is the MAIN reason.

Reasons for not continuing schooling may be as follows:

Still in school – has not completed high school and is still in the process of completing it. This applies to respondents between the ages of 16 and 18 years of age and adults continuing their education.

Had enough education – the education completed was sufficient for the respondent's needs.

Had to work/financial reasons – not able to afford to attend school on a full-time or continuous basis.

Wanted to work/wanted to learn a trade – refers to a hands-on approach where prerequisite did not require completion of high school.

Family reasons – ongoing commitment or responsibility to the family such as: illness in the family, having to work in family business, caring for children.

Did not like school/boredom – conflicts with teachers, curriculum not geared to respondent's wants, etc.

Did not do well in school – difficulty with course material, etc.

Personal illness or disability – respondent's own health problem.

School not available/not accessible – due to distance or political situation such as war. Also, for many seniors, schooling in the rural areas did not go beyond a certain level (for example grade 8). Continuing to higher levels meant boarding away from home. Therefore, the category "school not available" should also be interpreted as schooling beyond a certain level not available.

BEST COPY AVAILABLE

Don't know – probe if the respondent is unsure. Do not mark this category if the respondent refused to answer this question.

Other – ensure that the response does not fit into any of the listed categories. For example, "School wouldn't accept me because I was/am disabled" - would be reported under Personal Illness or Disability.

SECTION D. LABOUR FORCE INFORMATION

QUESTION B1 – What language did you first speak as a child?

Do NOT read categories to the respondent. Reading aloud the first two categories - English and French - may result in an under-reporting of other languages.

Indicate the first language spoken by the respondent in early childhood.

If the respondent spoke more that one language, mark more than one category only if those languages were used equally.

NOTE: If English (French on the French questionnaire) was checked along with other languages, follow the skip to Question B12.

QUESTION B2 – How would you rate your current ability to speak that language?

QUESTION B3 – How would you rate your current ability to understand that language?

QUESTION B4 – How would you rate your current reading skills in that language?

QUESTION B5 – How would you rate your current writing skills in that language?

Enter the code for the language reported in B1 in the boxes under "First Language". If more than one language was reported in B1 alter the question as appropriate, (i.e., insert the name of each language) and enter the code for the second language under "Second Language".

Ask Questions B2 to B5 for each of these languages. Accept the respondent's perception of "Poor", "Fair", "Good" and "Very Good".

QUESTION B6 – How old were you when you first started to learn English?

Mark one only. If the respondent indicates that he/she does not speak English mark the "Does not speak English" category. This may be evident if an interpreter is assisting with the interview. Note that it is possible that a respondent spoke English in childhood but no longer speaks English because of a change in language environment.

QUESTION B7 – When you were growing up, what language or languages were usually spoken in your home? (Accept multiple responses only if languages were spoken EQUALLY.)

This is not necessarily the language spoken most often by the respondent but the language spoken most often by household members.

QUESTION B8 – What language did you first learn to read and write?

If the language is not listed, mark the "Other" circle and enter the language in the space provided.

BEST COPY AVAILABLE

If the respondent reports learning to read and write two languages at the same time try to determine which was most frequently used and mark it.

QUESTION B13 – What languages (including English) do you speak well enough to conduct a conversation?

Accept the respondent's own interpretation of ability to speak well enough to conduct a conversation.

Follow the "Go to" to question B18 if the respondent speaks only one language.

QUESTION B14 – What language do you speak most often at home?

QUESTION B15 – What language do you speak most often at work or school?

QUESTION B16 – What language do you speak most often during leisure activities?

Mark only one. Even though a respondent may indicate more than one, ask him/her to indicate the main one.

This question refers to the language spoken **most often by the respondent**. It should not be interpreted to mean the language spoken most often by the family as a whole. It is quite possible that different family members speak different languages (e.g. Chinese children may speak English most often while the parents speak mostly Chinese).

If the respondent speaks two languages **equally at home**, report both.

If the respondent lives alone, report the language which is used on a daily basis.

Persons who first spoke an Indian language (from India) should not report "Indian" but rather Hindi, Urdu, Punjabi or some other language as applicable. Persons who speak an "Amerindian" language (North American Indian) should report Cree, Ojibway or some other Amerindian language as applicable.

QUESTION B17 – In which language can you express yourself most easily?

Make sure the respondent indicates the one they are most comfortable with for all activities - speaking, writing, reading, etc.

Those respondents who indicate that they are more comfortable in the other official language, must be asked if they wish to do the interview in the OTHER official language. If they indicate YES and you are unable to conduct the interview in that language, indicate that your supervisor will try to arrange an interview in that language and terminate the interview for now.

Inform your supervisor immediately.

QUESTION B18 – To which ethnic or cultural group did your ancestors belong?

Ethnic or cultural group refers to the respondent's "roots" and should not be confused with citizenship or nationality. Canadians belong to many ethnic or cultural groups such as Inuit, North American Indian, Métis, French, English, Irish, Scottish, German, Italian, etc.

Report the ethnic or cultural group rather than the language, for example, Austrian rather than German; Haitian rather than French; Scottish rather than English.

Mark all applicable answer categories. For example, if the respondent's parents were English and Irish, you would mark both of these categories.

If the respondent insists on answering "Canadian" mark the Canadian category and:

(a) PROBE by asking if his/her parents are "French-Canadian", "English-Canadian", "Irish-Canadian", etc. OR ask where his/her ancestors came from.

(b) if he/she answers that his/her parents or ancestors are "French-Canadian", mark the response category corresponding to the "French" prefix.

If the respondent replies that he/she is "Canadian" and after probing specifies that his/her parents are "Lebanese-Canadian", enter this information in "Other", since the prefix "Lebanese" is not a response category.

If the respondent **continues to insist** that his/her parents or ancestors are Canadian, accept that response.

SECTION J. HOUSEHOLD INFORMATION

QUESTIONS C1 TO C5

QUESTION C1 – The next few questions are about your mother (female guardian). **Can you answer some questions about her?**

QUESTION C2 – Was your mother (female guardian) **born in Canada?**

QUESTION C3 – Did your mother (female guardian) **immigrate to Canada?**

QUESTION C4 – Was your mother (female guardian) **under the age of 16 when she immigrated to Canada?**

QUESTION C5 – What was the highest level of schooling that your mother (female guardian) **ever completed?**

Questions C1 to C5 refer to the respondent's mother or female guardian. A female guardian can be a stepmother, a family member or a court-appointed guardian. These five questions are repeated in Questions C7 to C11 but refer to the respondent's father or male guardian. A male guardian can be a stepfather, a family member or a court-appointed guardian.

Mark "YES" for question C3 if, at any time, the respondent's parent or guardian immigrated to Canada. The parent or guardian may have immigrated to Canada at some time but may no longer be living in Canada.

For Question C5, you should refer to question A5 for explanation of schooling categories.

QUESTION C6 – Did/has your mother (female guardian) ever worked at a job or business?

Work means any activity carried out by the respondent's mother (female guardian) for pay or profit and includes "payment in kind" (payment in goods or services rather than money).

NOTE: Working at a job or business includes self-employed work and unpaid family work (i.e., activities which contribute directly to the operation of the business, farm or professional practice owned or operated by a related member of the same household).

BEST COPY AVAILABLE

QUESTION C7 – The next few questions are about your father (male guardian). **Can you answer some questions about him?**

QUESTION C8 – Was your father (male guardian) **born in Canada?**

QUESTION C9 – Did your father (male guardian) **immigrate to Canada?**

QUESTION C10 – Was your father (male guardian) **under the age of 16 when he immigrated to Canada?**

QUESTION C11 – What was the highest level of schooling that your father (male guardian) **ever completed?**

Questions C7 to C11 refer to the respondent's father or male guardian. A male guardian can be a stepfather, a family member or a court-appointed guardian.

For instructions on Questions C7 to C11, refer to Questions C1 to C4.

For Question C11, refer to Question A5 for an explanation of schooling categories.

QUESTION C12 – For whom did/does your father (male guardian) work?

If the respondent's father is not currently working then the respondent should report his main job while he was working. If the respondent's father has had more than one job, the respondent should report the main job. The main job is based on the respondent's perception.

If the father never worked, mark the "Never worked" circle.

QUESTION C13 – What kind of business, industry or service is/was this?

At least two words are generally required to provide an accurate description of the type of business, industry or service. For example, road construction, building construction, coal mine, iron mine, wholesale grocer, retail grocery store. This question refers to the same "main job" as reported in Question C12.

QUESTION C14 – What kind of work was/is your father (male guardian) **doing?**

The answer to this question should tell clearly and specifically the kind of work or nature of duties performed by the respondent's father (male guardian) at his main job or business. For example, office clerk, sales clerk, shipping clerk, factory worker, construction foreman, construction labourer.

SECTION F. ADULT EDUCATION AND TRAINING

QUESTION D1 – What is your current work situation? Are you...

Employed – Refer to "working at a job or business" below for a definition of **employed**.

Other includes the following:

- Maternity/paternity leave. This includes both people who are receiving maternity benefits from UIC and people who are on maternity/paternity leave without pay provided that they have a set date to return to work and that their employer has guaranteed them a job.
- Long-term illness or disability,

BEST COPY AVAILABLE

- On strike,
- On layoff,
- On sabbatical,
- Doing volunteer work,
- People who are living on investment income.

Before marking this category, probe to ensure that the answer really belongs here. Answers such as "taking it easy" may turn out to be situations that should be marked "unemployed" or "retired".

QUESTION D2 – Did you work at a job or business at any time in the past 12 months (regardless of the number of hours per week)?

"**Working at a job or business**" means any activity carried out by the respondent for pay or profit and includes "payment in kind" (payment in goods or services rather than money).

INCLUDE: Time spent in "on-the-job" training if the person is receiving pay from the employer in exchange for the goods or services produced in the course of training (e.g. apprentices, articling students of law, interns, etc.).

Time spent by an unpaid respondent in any activity that contributes directly to the operation of the business, farm or professional practice owned or operated by a member of the family.

EXCLUDE: Time spent in "on-the-job" training when the financial compensation received is not related to the goods or services produced, but is simply an allowance to enable the person to take the training (e.g. student nurses, armed forces reserves).

Time spent in government-sponsored adult training programs (e.g. Canada Manpower Training Program or other programs designed to enable adults to qualify for jobs) if this training is conducted solely in a classroom setting and the person is not working at a job or business.

Volunteer work without pay for organizations such as the Red Cross, United Way, local hospital, etc.

For **Self-employed** respondents:

INCLUDE: Time spent actively looking for potential clients, preparing estimates, quotes or tenders, etc.

Time spent operating a business, professional practice, or farm even if no sales were made, no professional services were rendered or nothing was actually produced.

Time spent on activities related to establishing a new business, farm or professional practice.

Time spent by a person who owns and manages his/her business or farm even though he/she is physically unable to do the work.

QUESTION D5 – Did you work mostly full-time or part-time?

Full time – defined as 30 hours or more of work per week.

Part time – defined as 29 hours or less of work per week.

If respondent worked full-time and part-time at the same time, mark "full-time".

If the respondent worked 30 weeks part-time and 15 weeks full-time, mark "part-time".

QUESTION D7 – For whom did you work the most hours in the past 12 months?

If the respondent is not currently working then the respondent should report the main job while he or she was working. If the respondent has had more than one job in the past 12 months then the main job should be reported. The main job is based on the respondent's perception.

QUESTION D8 – What kind of business, industry or service was this?

At least two words are generally required to provide an accurate description of the type of business, industry or service. For example, road construction, building construction, coal mine, iron mine, wholesale grocer, retail grocery store. This question refers to the same "main job" as reported in Question D7.

QUESTION D9 – What kind of work were you doing at this job?

The answer to this question should tell clearly and specifically the kind of work or nature of duties performed by the respondent's father (male guardian) at his main job or business. For example, office clerk, sales clerk, shipping clerk, factory worker, construction foreman, construction labourer.

QUESTION D10 – In total, about how many persons are employed by this business at all locations in Canada?

Some respondents may find it difficult to say exactly who they work for. For example: Is an interviewer for Statistics Canada a Regional Office employee, a Statistics Canada employee or a federal government employee? Emphasize that we are looking for employees **at all locations in Canada** and accept the respondent's perception of the organization he/she works for.

QUESTION D12 – What type of job was this? Was or is this job a...

A permanent job or work contract of unlimited duration is one in which there was no indication that the job would end at some definite point in time (e.g., hired permanently with no specified term).

A temporary job or work contract of limited duration is one in which there was a definite indication that the job would terminate at some specified point in time (e.g., hired for a six month term).

QUESTION D13 – How many hours per week did you usually work at this job?

In some jobs the number of hours worked each week may vary depending on the time of year. In these cases, try to determine the average over the year.

QUESTION D14 – During the past 12 months, how many weeks did you work at all jobs including time off for vacation, maternity leave, illness, strikes and lockouts?

The entry must not be more than 52 weeks. Time on Worker's Compensation is NOT to be included as weeks worked. Include maternity/paternity leave without pay as weeks worked provided the respondent has a firm job guarantee at the end of the leave period.

QUESTION D17 &

QUESTION D21 - During the past 12 months, for how many weeks were you without work and NOT looking for work?

The entry must not be more than 52 weeks.

If the respondent indicates that the time spent without work and not looking for work was less than one week, mark "01".

SECTION E. READING AND WRITING AT WORK AND LOOKING FOR WORK

This section of the Questionnaire is for those respondents who worked at any time during the last 52 weeks. They may not have worked for the full 52 weeks, or full-time.

QUESTION E1 – The following questions refer to the job at which you worked the most hours in the last 12 months.

How often do/did you read or use information from each of the following as part of your main job? Would you say every day, a few times a week, once a week, less than once a week, rarely or never?

"Every day" refers to every day that the respondent works. However, if the respondent works less than 5 days a week the "every day" category is not applicable. For respondents who do not work 5 days a week mark "**a few times a week**".

Reading material need not be written on paper. The respondent may read on a computer screen or a television screen.

QUESTION E2 – How often do you write or fill out each of the following as part of your main job? Would you say every day, a few times a week, once a week, less than once a week, or never?

Refer to E1 above for definitions of "every day" and "a few times a week".

Include writing/composition on a keyboard (e.g., typewriter, computer keyboard, etc.).

SECTION F. ADULT EDUCATION AND TRAINING

QUESTION F1 – The following questions will deal with any education or training which you may have taken in the past 12 months.

During the past 12 months, that is, since August 1993, did you receive any training or education including courses, private lessons, correspondence courses, workshops, on-the-job training, apprenticeship training, arts, crafts, recreation courses or any other training or education?

BEST COPY AVAILABLE

Ensure that you read the entire question to respondents, even if they say "yes" before you have finished. By mentioning a variety of different types of training or education, respondents will be prompted to remember all the training they have received throughout the year.

Conferences, seminars and workshops whose primary activity is considered as training or education, are also included.

QUESTION F3: What were the names (titles) of these courses or the programs associated with these courses?

Note that this question asks about courses or programs. A program is a collection of courses that leads to a specific degree, diploma or certification. Be sure to emphasize the word "programs" when you read this question. If the respondent is a student enroled in a program (e.g., Business Administration, Childcare, Computer Sciences) at a college or university, enter the name of the program under "First mention" and any other non-program courses in the other columns.

The interviewer box after this item reminds the interviewer to:

- list the 3 most recent courses if the respondent took more than 3 in the last 12 months; and
- prompt the respondent for all courses taken if fewer than 3 are mentioned.

QUESTION F4 – Now I would like to ask you about (insert name of course or program). Was this training or education financially supported by... Yourself or your family, an employer, the government, a union or professional organization?

You must mark either "Yes" or "No" for each BOLD FACE category.

If the respondent does not know who was supporting the training or says that there were no fees involved, mark "Yes" for "No fees" or "Don't know". These categories are not to be read to the respondent. If the respondent does not mention either of these categories, leave these circles blank.

If the respondent is self-employed, and the course was paid through the business, mark "**employer**" not "**yourself or your family**".

Mark "**the government**" if the respondent's training was being supported by any level of government, federal, provincial, or municipal. Welfare training programs are included in this category as all three levels of the government can be involved.

Examples of "**Anyone else**" are Lion's Club and Club and Medical Association.

Examples of support could be giving time off, paying for transportation, course material or tuition.

- When respondents take training during usual working hours and receive all or part of their salary, consider this "time off". Whether it is considered "time off" (i.e., employer support) when employers provide flexible hours or authorized vacation leave for employees to take courses, depends on the respondent's perception.
- Transportation includes plane fare, bus fare, train fare, vehicle mileage allowance, vehicle rental, etc.
- Course materials include text books, writing materials, protective clothing for use during the course, tools, etc.

Other examples of support include providing facilities, accommodation, an allowance, etc. An allowance is an amount of money allowed in a fixed period of time (by day, week, month, year) to purchase meals, dry cleaning, incidentals, etc. When in doubt, accept the respondent's perception of "financial support".

QUESTION F5 – Were you taking this training or education towards....

A university degree/diploma/certificate? A university is a degree/diploma/certificate granting institution which usually offers programs in at least the arts and sciences. Admission to university is usually dependent upon graduating from secondary school.

A college diploma/certificate? Include programs taken at colleges of applied arts and technology or CAATS (in Ontario), collèges classiques and CEGEPs (in Québec), and those programs which normally **require successful completion of secondary school.**

Technical institutes which offer diplomas and certificates at the college level include: institutes of agricultural technology, medical technology, Northern Alberta Institute of Technology (NAIT), Southern Alberta Institute of Technology (SAIT), British Columbia Institute of Technology (BCIT), Ryerson, and Eastern Ontario Institute of Technology (EOIT). Some, like Ryerson offer university degree programs as well and these should be reported as "university degree/diploma/certificate" if the respondent is enroled at the university level.

Nursing schools are considered colleges for the purposes of this question. Respondents taking a Bachelor of Nursing Science should report it as "university degree/diploma/certificate".

Teachers' College may be either an independent institution or a college associated with a university. An independent teachers' college usually awards a certificate or diploma. If the student is enroled in such a program, then it should be considered as "college". However, if a teacher's college is affiliated with a university that awards a degree in education and the student concerned is enroled in its program, then the certificate should be considered "university".

A trade-vocational diploma/certificate? A trade-vocational school is a public educational institution which offers courses to prepare people for employment in a specific occupation. Although entrance requirements vary, they frequently **require less than a high school diploma for admission.**

Vocational schools offer diplomas and certificates at the trade level. This term is used to classify skill courses that prepare trainees for occupations not at the professional or semi-professional levels. The emphasis is on manipulative skills with varying degrees of complexity and the performance of established procedures and techniques.

Examples:

- mechanics; welding; machinists; metal work; construction trades, such as, carpentry, electricity and masonry; secretarial typing; accounting; nursing assistant, dental assistant; cooks; medical laboratory technicians; waitresses; barbers, hairdressers, etc.

Trade-vocational and technical schools are often used synonymously and technical and trades training varies between and within provinces. It is offered in both public and private institutions such as community colleges, institutes of technology, trade schools and business colleges. Note that "trade-vocational diploma/certificate" at vocational schools are considered secondary education whereas college or technical institutes offer diplomas at the post-secondary level and are included in college or university programs for this question.

Private or commercial schools, although they are profit oriented (i.e., for monetary gain), are licensed by the province, and are engaged in providing professional and vocational training; for example, hairdressing, data processing, dancing, truck driving and the operation of heavy equipment, languages, dressmaking, secretarial, etc.

An apprenticeship certificate? This is a systematic program that combines on-the-job training (by observation and practice) with shorter periods of related in-class training (by attending technical courses where the more theoretical aspects are taught) to produce a fully qualified journeyman tradesperson. The employer undertakes, by contract, to employ and train an apprentice under the supervision of a qualified journeyman. In Quebec it is possible to receive apprenticeship training without related in-class training.

An elementary or secondary school diploma? These may be taken at a private or a public school.

Professional or Career upgrading? These could be courses, workshops, seminars or conferences offered by the employer or private businesses. They may be paid for by the employer or by the respondent.

QUESTION F6 – Was this training or education given by....

a university or other higher education establishment – "Other higher education establishments include Community Colleges.

a further education college – These are special institutions devoted to adult education. Examples of types of programs offered are job training, career development programs or high school upgrading.

a commercial organization (for example a private training provider) - These organizations specialize in offering courses and seminars to businesses for their employees on a wide range of topics (e.g., Stress Management, Time Management, Effective People Skills, etc.).

a producer or supplier of equipment – These courses are provided by a producer or supplier to the purchaser of a piece of equipment to show the purchasers employees how to use the equipment.

a non profit organization such as an employer association, voluntary organization or a trade union – Examples are St. John's Ambulance first aid courses or union workplace safety courses.

an employer or "Parent" company – Large companies such as MacDonald's, Black's Camera often run courses for store managers or franchise owners to help them produce a better product or run a more efficient business.

Mark this category only if the course was developed by the respondent's employer or a contractor for the employer. If the course was paid for by the employer but given by a private training provider or commercial organization mark "a commercial organization" above.

Other provider – This category is not to be read to the respondent. Mark "Yes" only if the respondent's answer does not fit in any of the above categories. Otherwise leave this item blank.

QUESTION F8 – For how many weeks did this training or education last?

If less than one week, enter "01". Respondents should also include the period devoted to exam writing.

When the respondent replies in months, calculate the weeks using 13 weeks for every 3 months. That is, 3 months is 13 weeks, 6 months is 26 weeks, 9 months is 39 weeks and 12 months, 52 weeks. If the respondent gives an answer like four months, add 4 weeks and 13 weeks to get the correct response of 17 weeks.

Some school systems operate by semester. If the respondent reported being enroled in a program in question F3, enter the number of weeks the respondent was in school for that program in the last 12 months. So if a respondent took 4 courses, over 2 semesters and each semester lasts 13 weeks, enter 26. Since semesters vary by province and by educational institution, probe to find out what is the duration of the semester.

QUESTION F10 – On average, how many hours per day was it?

If the number of hours vary each day, and the respondent is unable to provide an average, ask for the number of hours per week and divide by the value reported in Question F9.

QUESTION F11 – What was the main reason you took this training or education? Was it for....career or job-related purposes? ... personal interest? .... other?

"For career or job-related purposes" means the purpose of the training/education is to learn new skills for a current or future job; increase earnings; improve job opportunities in current or another field, improve opportunities for advancement/ promotion, etc.

"For personal interest" means programs that are taken for a reason not related to an individual's job or career. This would include courses or programs designed to improve social skills and courses taken for reasons of personal development.

QUESTION F12 – To what extent are you using the skills or knowledge acquired in this training or education at work? To a great extent... Somewhat.... Very little... Not at all.

When respondents are full-time students, indicate the "Not at all" response category. Similarly, if respondents are currently not employed, and they took employer-supported training in the past and used it on the job, mark the response "Not at all" because they are not using it at the time of the survey.

This may seem to be a strange question when the program is being supported by the employer. However, some employers support programs which may have nothing to do with the employee's current job, but may help them in their career aspirations. For example, a machine operator in a chemical factory may be supported by the employer as he/she works towards a degree in English Literature and upon successful completion hopes to become a technical writer for the employer. At the time of the survey the respondent may report that he/she is not using the knowledge and skills of the degree program as a machine operator.

QUESTION F14 – Was this training or education provided through...

Indicate "Yes" or "No" for each category.

Classroom instruction, seminars or workshops?

In this situation there is an instructor or teacher who leads a group in its learning, face to face (as opposed to through electronic media). Classroom instruction can take place anywhere, not only at a school.

Educational software?

"Educational software" is a computer program designed to teach any number of subject matter. For example, software packages that teach mathematics, writing skills and geography. Other examples are the tutorial packages included with the purchase of computer software programs such as Wordperfect, Supercalc, etc.

The respondent could have used a computer exclusively to learn, as in computer-based tutorials, or could have used the computer in conjunction with an instructor. Mark "Yes" even if the computer was not the main method of instruction.

Exclude the use of software (such as Wordperfect) when it is used to type assignments or find books in the library.

Radio or TV broadcasting

Audio/video cassettes, tapes or disks?

Correspondence courses are sometimes conducted through various electronic media. This type of training is also referred to as "Distance education".

"Distance education" is any form of education in which the teachers and students are not at the same place. Students usually study from their own home, on a part-time basis. Instruction is received via mail, in the form of reading assignments or exercises and/or electronic media (such as television, audio-visual tapes, audio cassettes, etc). In British Columbia, programs taken through the Open Learning Institute qualify as distance education.

NOTE: The student must have been registered in the program to be included.

If the respondent recorded a TV or radio broadcast for later use, mark "Radio or TV broadcasting", not "Audio/video cassettes, tapes or disks".

Mark "Yes" for these categories even if they were not the principal means of instruction.

Reading materials?

Mark this category if the course taken included use of any type of reading material, for example, textbooks, handouts.

On-the-job training?

On-the-job training" is instruction in the work environment.

Include in this category, any informal on the job training as well as apprenticeships, internships or articling training.

In trades, apprentices work under the supervision of a qualified tradesperson and learn the principles, skills, tools and materials of the trade. An apprentice enters into a contractual agreement with an employer to learn a trade and has registered this contract with the provincial ministry responsible for apprenticeship.

"Articling" refers to the period of training (usually 12 months) in which a graduate from a Bachelor of Law degree program practices law under the supervision of a solicitor, as part of the requirements for obtaining his licence to practice law in all provinces. Similarly, accountants also article.

"Internship" is the period of training required by a medical school graduate in which the graduate performs the duties of assistant physician or surgeon.

Other methods?

This category is not to be read to the respondent. Mark "Yes" only if the respondent's answer does not fit in any of the above categories, for example, one-on-one instruction with a tutor. Otherwise leave this item blank.

INTERVIEWER

This interviewer Item reminds the interviewer to return to question 4 if the respondent reported more than one course in Item F3.

QUESTION F17 – Since August 1993, was there any other training that you WANTED to take but did not, such as hobby, recreational or interest courses?

"Hobby or recreational courses" are courses taken for the purposes of learning a hobby; physical, social, or psychological development; or personal interest in a particular subject-matter.

SECTION G. READING AND WRITING GENERAL

The questions in this section apply to reading and writing and activities done in any language not just English or French.

QUESTION G1 – The next few questions deal with reading and writing in your daily life, excluding work or school.

I am going to read you a list of activities. Please tell me if you do each of them daily, weekly, every month, several times a year or never? How often do you...

Count the activity even if it was not done in English or French.

For **c) attend or take part in a sporting event**, include participation or attendance at organized team sports only. Exclude recreational swimming, cycling, etc.

For **e) participate in volunteer or community organizations**, include any type of volunteer work such as coaching a little league team, sitting on a Board of Directors of a non-profit organization, helping raise money for a school, church or other organization.

QUESTION G3 – Which of the following activities have you ever done in a language other than French or English?

The purpose of this question is to determine if any of the activities reported in G1 were done in a language other than English (or French). The activities are the same as those in question G1.

BEST COPY AVAILABLE

QUESTION G6 - Which of the following materials do you currently have in your home? daily newspaper, weekly magazine, more than 25 books....

Mark "Yes" or "No" for each type of material.

Mark "Yes" even if the material in a language other than English or French.

QUESTION G11 - Sometimes people need help from family members or friends to read and write in English. How often do you need help from other with...

We are interested in help received from family and friends; do not include co-workers or professionals such as one's accountant unless they are considered family or friends.

QUESTION G16 - Did you ever have...

a) **eye/visual trouble of the kind that is not corrected by glasses?**
b) **hearing problems?**
c) **a speech disability?**
d) **a learning disability?**
e) **any other disability or health problem of six months or more?**

If the respondent says "No", to each part of the question, go to question H1.

If the respondent says "Yes" to any part of the question, follow the arrow and ask questions G17 and if the response is "Yes" again, ask G18.

QUESTION G17 - Did you have this problem while you were in primary or secondary school?

QUESTION G18 - Do you have this problem now?

If the respondent says "No, to each part of question G17, go to question H1.

If the respondent says "Yes" to any part of the question, follow the arrow to G18.

SECTION D. LABOUR FORCE INFORMATION

QUESTION H1 - Are you the parent or guardian of any children aged 6 to 18 that are presently living with you?

Include 6 and 18 year olds. If the answer is "No", go to question J1.

QUESTION H3 - What is the highest grade of schooling that this child has completed?

For the **Elementary** and **Secondary** categories, enter the highest grade completed. For the Trade/Vocational and Post-Secondary categories, enter the number of years completed.

If necessary, refer to Question A5 for definition of these categories.

BEST COPY AVAILABLE

QUESTION H4 - How often would you say this child reads for pleasure? Would you say ... every day..... a few times a week.... several times a month....

Include any reading or attempt at reading by the child. Include reading done with the help of an adult provided the child wanted to do it.

Exclude - reading done by the parent to the child, (e.g., bedtime stories).
- reading done as part of homework assignments.

THE NEXT QUESTIONS WILL DEAL WITH YOUR ENTIRE HOUSEHOLD

QUESTION H7 - Could you please tell me if each of the following statements are true or false of your household?

a) There is a variety of books in your home..... b) There is a variety of magazines and other reading material in your home.... c) Your children often see you or your spouse reading ....

For this question as for the questions in Section G, the books or magazines need not be English or French. Mark "True" if books and magazines are available or being read in any language.

SECTION J. HOUSEHOLD INFORMATION

QUESTION J1 - Finally, I would like to get a bit of general household information.

From which of the following sources did you receive income in 1993? (Mark all that apply)

a) **Income from wages, salary or self-employment**

Include all earnings from tips, commissions, gratuities as well as all earnings from any type of casual work such as paper delivering, babysitting, house cleaning, etc. Any pay received as a member of a reserve unit is included as income.

b) **Income from governments such as Family Allowance, Unemployment Insurance or Social Assistance?**

Include:

Worker's Compensation income received for injuries sustained at work.

Family and Youth allowances received from federal and provincial programs.

Unemployment insurance total benefits received from UIC. Include benefits for sickness, maternity/paternity leave, work sharing and retraining.

Social assistance and income supplements are made to persons in need by federal, provincial or municipal governments. Benefits include assistance to mothers/fathers with dependent children, elderly individuals, the blind and the disabled. Include cash benefits covering basic needs (e.g. food, fuel, shelter, clothing). Also include income received from Worker's Compensation for injuries sustained at work.

c) **Canada or Quebec Pension Plan, or Old Age Security?**

BEST COPY AVAILABLE

These government pension plans also make payments for injuries and disabilities and to surviving spouses (Widow(er)'s benefits).

d) **Income from interest, dividends, investments or private pensions.**

Dividends are payments made to shareholders.

Private pensions are pensions received from a company that the respondent worked for.

Include any capital gains, income from RRSPs (RIFs) interest on bonds, dividends, cash dividends from life insurance policies, corporate stocks, royalties on books, oil wells, rental income.

e) **Income from any other sources such as alimony, scholarships, etc.**

Any other money/income NOT reported in items a), b), c) and d) above, such as maintenance payments and support payments from a former spouse, non-refundable scholarships, income from outside of Canada, severance pay, etc.

QUESTION - J2 What is the best estimate of your personal income in 1993 from all sources, including those just mentioned?

Report the total amount from all jobs before deductions for income tax, pension contributions, health insurance, Canada Savings Bonds, etc.

EXCLUDE: gambling gains, lottery winnings, money inherited during the year in a lump sum, income tax refunds, loan payments, lump sum settlements of insurance policies, rebates received from property taxes, pension contribution refunds, strike and sick pay from trade unions, etc.

If the respondent hesitates, remind them that this information is strictly confidential.

If the respondent reports income in a foreign currency, ask him to convert the amount into Canadian dollars (to the best of his/her ability) or indicate which currency. When doing edits on the questionnaire calculate the amount using the current bank rates.

QUESTION J3 - What is the best estimate of your personal income from only wages, salary or self-employment in 1993?

Include all earnings from tips, commissions, gratuities as well as all earnings from any type of casual work such as paper delivering, babysitting, house cleaning, etc. Any pay received as a member of a reserve unit is included as income.

7. INTERVIEWER - Record end time of BQ

Record the time you finished the Background Questionnaire, using the 24 hour clock.

PART II: CORE TASKS - SCORING

You will use this section of the questionnaire to score the Core Tasks Booklet.

8. INTERVIEWER - Language of Task Booklets

The respondent may have completed the Background Questionnaire in one official language but may prefer to complete the Core and Main Tasks in the other official language. Ensure that you indicate the language selected for the completion of the Core Tasks in this space.

If the respondent chooses to complete the Core Tasks in English he/she must complete the Main Tasks in the same language. If the respondent wishes to change languages after starting the Tasks Booklets you must re-administer the tasks previously completed.

9. INTERVIEWER - Record start time of Tasks

Record the time you start the Core Tasks Booklet, using the 24 hour clock.

At this point you will continue the interview using the Tasks Administration Guide. Once the respondent has completed the Core Tasks you will score his/her answers using Part II - Core Tasks Scoring starting at Item 1.

Scoring the Core Tasks Booklet

As soon as the respondent has finished the Core Tasks, you are to complete the Core Tasks - Scoring section of the questionnaire. Accurate scoring of these tasks by the interviewer is critical if the survey results are to be accurate. Ensure that you review the Core Tasks Booklet and this section of the questionnaire carefully before your first interview so that you know what to look for.

You will score the 6 tasks which make up the Core Tasks. If the respondent completes 2 or more items correctly in the Core Tasks Booklet (score value of "1"), you will continue the interview by administering the Main Tasks Booklet.

FIRST CIRCLE (CODE 1)

The first circle for each item on the scoring form represents the correct answer to the question. If you have marked at least 2 code 1 circles, you must ask the respondent to complete the Main Tasks Booklet.

SECOND CIRCLE (CODE 7) - ANY OTHER RESPONSE

Mark this circle if the respondent has made an incorrect entry or marking for the item in the Core Tasks Booklet.

BEST COPY AVAILAB

THIRD CIRCLE (CODE 0) - TASK REFUSED/NOT DONE

Mark this observation if:

(a) the respondent has verbalized a refusal. For example: "I don't want to do this." "I'm not going to do this."

(b) the respondent has not done the task for whatever reason: there is no entry for that question in the Core Tasks Booklet.

This circle cannot be marked if the respondent has attempted the task and made a mark in the booklet.

A respondent may attempt one or two tasks in the Booklet and then say, "I can't do any more. These are too difficult for me." Encourage the respondent to look at all six items. If the respondent maintains that the tasks are too difficult, mark the observations for the remaining core items as "Task Refused/not done" and terminate the interview.

Follow these guidelines when scoring the Core Tasks:

- Consider responses correct ONLY when they are identical or equivalent to the responses given for **circle 1** in Items 1 to 6 of Part II.

  So, for Item 1, the Unicef Ad, if the respondent circled "15 cents" you would mark Item 1 in Part II of the BQ as follows:

1.	**UNICEF Ad**
^{1}O	Underlines the sentence or the part of the sentence that contains "15 cents". *(circles or written response is also valid)*
^{7}O	Any other response.
^{0}O	Tasks refused/not done.

- If the question asks for a written response in a given space, but the respondent underlines, circles or enters the correct answer somewhere else on the page, they should receive a correct score of 1.
- Underlining or circling that extends beyond the start or end of a sentence will be considered correct ONLY when the underscore starts and/or ends on the same line that the sentence start or ends. If the underscore or circle extends to a new line the response will be considered incorrect.

Score carefully, your scoring may determine whether or not the respondent completed the Main Tasks Booklet. If the Main Tasks Booklet is completed unnecessarily the result will be higher survey costs.

A certain percentage of Core Tasks will be rescored by your Senior as a quality control measure but you will not be able to return to administer the Main Tasks Booklet if it was not administered due to a scoring error. If the Main Tasks Booklet is not completed, valuable information will be lost.

INTERVIEWER

If respondent has completed two or more of the 6 core tasks correctly (score values of 1) proceed with MAIN TASK BOOKLET.

Otherwise, thank the respondent and end the interview. In Interviewer Check Item 5, code the appropriate status for the CORE TASK BOOKLET and code the status for the MAIN TASK BOOKLET AS "Z". Record the END TIME in the space provided below.

This interviewer item is self explanatory. It reminds you how to proceed once you have scored the Core Tasks Booklet.

10. INTERVIEWER

If respondent failed CORE Tasks, enter the END TIME below. Otherwise, proceed with MAIN TASK BOOKLET and Enter END TIME upon completion.

This item is self explanatory. A reminder will appear at the end of instructions for each Main Task Booklet in the Tasks Administration guide. If the respondent does not finish the Main Tasks, you must remember to make this entry.

REASON FOR NON-INTERVIEW

This section is to be completed each time that the entries in Item 5, Final Status on the cover page of the questionnaire are not all Code X (Complete) or Code P (Partial). If any one box, either BQ, Core or Main contains a code other than X or P, this section must contain a detailed explanation of the reason for the non-interview. For example: "The respondent speaks English well enough to complete BQ but does not read or write English or French well enough to complete Core and Main Tasks."

Explanations for Refusal cases should include a summary of the interviewer's efforts and the respondent's comments. For example: "Called the respondent 3 times to schedule an appointment -kept saying he was too busy to be interviewed today. Finally said he had provided enough info for Stats Can surveys in the past and wasn't doing any more."

BEST COPY AVAILABLE

SECTION 4 - REVIEW EXERCISE

Circle the letter beside the correct answer or enter the correct response.

1. Marie was born in Germany to Canadian parents who registered her as Canadian at the time of birth so you would mark "Yes" for question A1 "Were you born in Canada"?
   - (a) True
   - (b) False
2. Richard completed his Secondary school program and is currently taking courses at a Community College towards his diploma in Computer Sciences. Which category would you mark in Question A8?
   - (a) Some Secondary.
   - (b) Probe to determine if he completed a general, academic, vocational or technical program in high school and mark the appropriate category.
   - (c) Completed non-university Post-Secondary.
3. In order for a person to be considered as having worked, he/she must have received a wage or salary.
   - (a) True
   - (b) False
4. In question D14, Francine indicates that she worked for 7 months and was on maternity leave for the rest of the year. Indicate which answer should be recorded.
   - (a) 7 months
   - (b) 28 weeks
   - (c) 52 weeks
   - (d) either (a) or (b) above

5. Pierre worked for 42 weeks and was on lay-off but did not look for work for the rest of the 12-month period. How many weeks would be recorded in Question D17?

   (a) 00 weeks

   (b) 10 weeks

   (c) 07 weeks

   (d) either (b) or (c) above

6. When you ask question F14 and read the term "Educational software" to the respondent, he asks you what that means. How would you explain this term?

   ____

   ____

   ____

   ____

   ____

7. In response to question F3, a 19 year old respondent tells you that he has taken 6 courses since January of 1994. What do you enter under "First, Second and Third mention for this question?

   a) Just enter the first 3 courses mentioned.

   b) Determine if the respondent is enroled in a program and if so, enter the name of the program under "First mention" and other courses under 2nd and 3rd mention.

   c) If the respondent is not enroled in a program, enter the 3 most recent courses.

   d) b and c above

8. If you have marked the first circle (code 1) in the Core Tasks - Scoring section of the questionnaire, this means that the respondents answer is ______________.(correct\incorrect?)

5 DESCRIPTION OF THE TASKS BOOKLETS

5.1 Introduction

All respondents will complete the Core Tasks Booklets and a pre-selected Main Tasks Booklet indicated on the label of the Background Questionnaire.

This section will provide you with an understanding of the skills needed by respondents to complete the pre-selected Task Booklets, and an appreciation of the problems which may be encountered by respondents with low literacy levels. While the majority of respondents will be able to complete the tasks in the Booklet with relative ease, some may find them difficult. The information in this section will help you to react to situations which may arise when you are administering the Tasks Booklets.

The tasks in the Booklet simulate those which people are exposed to in the course of their daily activities. Some respondents may say they are never called upon to do certain tasks that are included in the survey and see no need to do them now. If this situation arises, explain to the respondent that the survey is intended to provide a literacy profile of all Canadians and, while not all items will be specific to each individual, they combine to form a more complete profile of the literacy levels of Canadians. For this reason, you would appreciate their co-operation in completing the items.

There is no time limit for completing the Core or the Main Tasks. Allow the respondent all the time required. You may encounter cases where it is obvious that the respondent cannot do the task but is reluctant to give up. In such cases it is appropriate to suggest that the respondent try another question and return to the one which is causing difficulty later in the interview.

5.2 Interruptions during the Interview

If you are unable to conduct the entire interview for the Background Questionnaire at one time (e.g., the respondent must leave for an appointment, child care responsibilities, illness) you may return at some other time to complete the BQ and the Task Booklets. However, if the interview is interrupted during the Core or Main Tasks, you cannot return to continue the interview.

Respondents may take time out from the tasks to care for children, answer the phone or get a cup of coffee but if you must leave the respondent's home after some of the tasks have completed, consider the case as ended.

5.3 Administering the Core Tasks Booklet

The Core Tasks Booklet is to be administered once the Background Questionnaire is completed. The Main Tasks Booklet is to be administered if the respondent has correctly completed at least 2 items in the Core Tasks Booklet.

The introduction to the Tasks Booklets, the special instructions to the interviewer and instructions to be read to the respondent are found in the Tasks Administration Guide. There are 8 sections in the Tasks Administration Guide, 1 for the Core Tasks Booklet and 1 each for the 7 Main Tasks Booklets. The appropriate Main Tasks Booklet to use will be identified on the label of the Background Questionnaire.

The Tasks Administration Guides will provide you with:

- Special instructions (i.e., hand newspaper to respondent, limitation in assistance to respondent, etc.), and
- Instructions to be read by you to the respondent for each task contained in the Tasks Booklets.

When administering the Tasks Booklets, read all instructions at a moderate speed and allow the respondents as much time as they need or as much time as they request, to complete a particular item.

Once the respondent has completed all the items you will score the booklet using the Core Tasks - Scoring section of the Background Questionnaire.

If the respondent is showing any signs of nervousness, take a moment to establish a non-stressful environment. Comments such as, "That's fine" or "You're doing okay" may be appropriate at this point.

If the respondent is unwilling or unable to complete an item, ask if he would like to go to the next item. Respondents should be told that they may return to any skipped item in a section.

Proceed as follows to administer the Core Tasks.

a) Keep the Background Questionnaire open to Part II - Core Tasks Scoring.

b) Turn to the first page of the section labelled "Core" in the Tasks Administration Guide included with your training supplies.

- The box at the top of the page contains the following instructions:

> ***INTERVIEWER:***
>
> *Before you administer the tasks, record the R.O. Docket number on the Core Tasks Booklet; hand the respondent a pencil and the Core Tasks Booklet, then read the introduction.*

- Read the introduction that follows this box to the respondent.

c) Enter the respondent's language choice and log the start time for the Core Tasks in PART II of the BQ as instructed in the other interviewer box at the bottom of the first page of the Tasks Administration Guide.

d) Turn to page 2 of the Core Tasks section in the Tasks Administration Guide and read the instructions to the respondent.

e) Once the respondent has completed the Core Tasks, score them using Part II: Core Tasks -Scoring of the BQ. If at least 2 of the respondent's answers correspond to a Code 1 in the Core Tasks Scoring section, ask him\her to complete the Main Tasks. Otherwise, end the interview.

5.4 Administering the Main Tasks Booklets

There are seven different Main Tasks Booklet. Each one contains 3 of 7 question blocks in a different order in each booklet. This order of "spiralling" of questions has been done in order to avoid bias. By varying the order of the blocks of questions in the booklets it is hoped that the effects of respondent fatigue on survey results will be reduced.

The table below shows the question blocks that you will find in each Main Tasks booklet.

Booklet	Contains Blocks:
1	1, 2, 4
2	2, 3, 5
3	3, 4, 6
4	4, 5, 7
5	5, 6, 1
6	6, 7, 2
7	7, 1, 3

The page numbering system for each block begins with the Block number indicator and continues sequentially. For example, the first page in block 3 is page 3-1, the second 3-2, etc. Because of the spiralling of Blocks within booklets, the page sequence within booklets will seem very odd. In booklet 1, the page numbers go from page 2-8 to page 4-1 because block 3 is not contained in booklet 1. Booklet 7 starts with page 7-1 to 7-10. Page 7-10 is followed by page 1-1.

Each block contains about 15 questions. The respondent will have to use the newspaper to answer some of these questions. If a respondent finds a question difficult encourage him to try the next one. The questions within each block increase in difficulty. If a respondent is getting to the end of a block and wishes to end the interview because the questions are too difficult, as a last resort, suggest that he try the questions in the next block.

Scoring of the Main Tasks Booklet items will be done in Head Office. The interviewer will simply direct the interview using the appropriate section of the Tasks Administration Guide.

SECTION 5 - REVIEW EXERCISE

Circle the letter beside the correct answer.

1. If you must leave the respondent's home you can return to complete interview provided that the interruption occurred

   a) before the Main Tasks were begun

   b) before the Core Tasks were begun

   c) during the interview for the BQ

   d) b) and c) above

2. The scoring for the Core Tasks Booklet is to be done:

   (a) after each item is attempted by the respondent

   (b) after you have left the respondent's home

   (c) immediately after the respondent has completed the Core Tasks

3. You may help a respondent to complete the Core Tasks Booklet.

   (a) True

   (b) False

4. Before beginning the Main Tasks Booklet, the respondent must:

   (a) complete all items in the Core Tasks Booklet

   (b) correctly complete two or more items in the Core Tasks Booklet

   (c) have no more than 3 items incorrect in the Core Tasks Booklet

5. The entry in Interviewer Item 10 in Part II of the Background Questionnaire could represent:

   a) the end time for the Core Tasks if the respondent failed the Core Tasks.

   b) the end time for the Main Tasks if the respondent completed the Main Tasks.

   c) the end time for the Main Tasks if the respondent refused to continue the Main Tasks at any time.

d) b) and c) only
e) All of the above.

6. If a respondent is taking a long time to read and write the answer to a question, you should:

   (a) Tell him that time is up.
   (b) Allow him to take as much time as needed to complete the item.
   (c) Tell him he should move on to the next item.
   (d) None of the above.

7. If a respondent is taking a long time to complete a question and you sense that he does not want to admit that he does not know the answer you should:

   (a) Tell him that time is up.
   (b) Allow him to take as much time as needed to complete the item.
   (c) Suggest that he try another item.
   (d) None of the above.

6 EDIT STEPS

To ensure that all forms contain accurate information, you must edit each Background Questionnaire.

6.1 Edits to the Background Questionnaire

Head Office will be required to process the data from this survey within very tight time constraints. For this reason it is critical that each questionnaire that you return be completed as accurately as possible. While the respondent is completing the Main Tasks, you must perform the edits listed below on his/her Background Questionnaire. If the respondent fails the Core Tasks, try nonetheless to take a minute before you leave his home to edit the BQ.

A. EDITS TO BE DONE WHILE RESPONDENT COMPLETES MAIN TASKS

Interviewer Item 4 - Assistance provided

Indicate if the respondent needed assistance for the completion of the Questionnaire.

Interviewer Items 6 and 7 - Start and end times BQ

Ensure that the start and end times for the Background questionnaire were entered in these Interviewer items on pages 3 and 22 of the BQ.

Flows (Go to...)

Ensure that the flows throughout the questionnaire have been correctly followed. Each time you have marked an answer that is followed by a "go to" instruction, circle the instruction to ensure that you have seen it.

Questions B2 to B5

If the answer to B1 is a language other than English (or French on the French questionnaire) there must be entries in questions B2 to B5. If this is the case, ensure that the proper language code(s) have been entered in boxes in questions B2 to B5.

Questions B13 to B17

Ensure that the language(s) reported in B14 to B17 have also been reported in B13.

Other - specify

Review all questions that have an entry in the "Other - specify" category and ensure that the entry you have made does not actually belong in one of the listed categories.

General edits

Ensure that:

- all written entries are complete, correct and legible;
- an "X" has been marked in the appropriate circle;
- the time entries are recorded using the 24-hour clock.

B. FINAL ENTRIES - TO BE DONE AFTER MAIN TASKS COMPLETED

Interviewer Item 8 - Language of Tasks Booklets

Ensure that this item has been marked.

Interviewer Items 9 and 10 - Start and End Time Task Booklets

Ensure that the start and end times for the Tasks Booklets have been entered in these Interviewer items in Part II of the BQ.

Interviewer Item 2 - Total Number of Calls

Verify that the total number of calls from the Record of Calls and Appointments is recorded.

Interviewer Item 3 - Interviewer's Name

Ensure that you have signed your name after editing and recording the final status of the all documents.

Interviewer Item 5 - Final Status

Ensure that a Final Status code has been recorded for all 3 survey documents: the Background Questionnaire and the Core and Main Task Booklets.

Transcribe these codes into the pre-printed Final Status boxes in the upper right-hand corner of the respondent's envelope.

Non-interview section

If you have not entered Code X or P for any of the documents in Item 5 - Final Status, explain the reason for the non-interview in the Non-Interview section on the last page of the questionnaire.

AFTER THE INTERVIEW, ENSURE THAT THE R.O. AND DOCKET NUMBER HAVE BEEN TRANSCRIBED ONTO EACH DOCUMENT AND THAT ALL DOCUMENTS ARE RETURNED TO THE RESPONDENT'S ENVELOPE WHETHER THEY WERE COMPLETED OR NOT.

NOTE: **Refer to your Final Status Code Sheet for a summary of the codes for each document. Whenever possible the standard LFS codes have been used.**

SECTION 6 - REVIEW EXERCISE

1. Why is thorough interviewer editing so important for this survey?

---

2. The language selected by the respondent for the completion of the Tasks Booklets does not need to be indicated in Interviewer Item 8 because the language of the interview is already noted in Item 1 on the cover page of the Background Questionnaire.

   (a) True

   (b) False

3. The total number of calls must be recorded on the cover page of the Background Questionnaire.

   (a) True

   (b) False

4. The completed Tasks Booklets and Newspaper should be kept together in the respondent's envelope with the Background Questionnaire.

   (a) True

   (b) False

7 FINAL STATUS

7.1 General

Each questionnaire must be assigned a final status in Item 5 on the cover page of the Background Questionnaire. If you encounter a refusal situation or the selected respondent will not be home during the entire survey period, enter the appropriate code on the Background Questionnaire and the Core and Main Tasks Booklets. Once all documents in the envelope have reached final status, transcribe the codes from Item 5 of the Background Questionnaire to the Final Status box on the respondent's envelope.

The final status codes are defined below. Some codes may seem inappropriate for certain documents. For example you would not normally use Code E, Interview prevented due to blindness or visual impairment on the Background Questionnaire as this questionnaire is administered verbally. However, if the respondent refuses to participate in the survey because he/she is blind, enter Code E on the survey documents for this respondent, not code R (Refusal).

7.2 Final Status Codes

Code X: Completed Questionnaire

Background Questionnaire — Contains no more than 5 unanswered questions.

Core and Main Tasks Booklets — There is evidence that the respondent attempted each question. Evidence means that the respondent made a mark on the page . This mark need not be the correct answer. It could be a question mark for example. This is nonetheless evidence that the respondent attempted the question.

Code P: Partial Questionnaire

Background Questionnaire — More than 5 applicable questions were not answered but most of the applicable questions were answered.

If the respondent provides enough information for a "Partial" BQ, attempt to have him/her complete the Tasks Booklets. If the Background Questionnaire does not meet the requirements for a Partial (i.e., it is a Code R, D, G, H, etc.), end the interview.

Core Tasks Booklets — There is no mark in the booklet for 1 or more questions.

If the respondent completed only 2 items on the Core Tasks Booklet and made no entry whatsoever for the other items, the Final Status of the Core Tasks Booklet would be P. If the two answered questions are correct, the respondent should still be asked to complete the Main Tasks.

Main Tasks Booklet - There is evidence that the respondent attempted at least one question on each answer page.

Code R: Refusal

The respondent refuses to participate in the survey or begins the survey and refuses to continue before the document contains enough information to qualify as a partial. Enter Code R for the document where the refusal occurred and for all subsequent documents.

EXAMPLE: The respondent completed the Background Questionnaire but refused to continue after completing only 2 items in the Core Tasks Booklet. The documents in his envelope would be coded as follows: BQ = X, Core = P, Main = R.

Code T: Temporarily Absent

You have contacted the household but have been informed that the selected respondent will be absent during the entire survey period. Also use Code T when you have reliable information that no household member will be at home during the survey period. For example:

- a neighbour has told you that all members of the household are on vacation in Florida from September 1st to 30th.
- the selected respondent is away on a business trip and will not be back for three weeks.

In all cases specify the reason and, if possible, the duration of the absence in the "Reason for Non-Interview" section of the questionnaire.

Code N: No one at home

This code applies to the dwelling in which the respondent now lives. Use Code N if, after numerous attempts, you have been unable to contact a member of the household. Also use Code N if you are unable to meet with the respondent. If, however, you are told that the respondent will not be home for the entire survey period, use Code T.

Use Code N also if, on your final visit to a dwelling, you believe that someone is inside but no one answers the door, or you actually see an adult who does not answer the door. DO NOT treat this as a refusal.

Code K: Interview prevented by death, sickness or other unusual circumstances related to the household or to the selected respondent.

Use Code K when unforeseen circumstances, for example a serious but temporary illness that cannot be categorized under disabilities (Codes A to H), death in the household or any other unusual situation prevent the interview from taking place. Other unusual situations include: house in quarantine, household recovering from fire or flood.

Use your best judgement, be tactful and ready to recognize if your presence in the household is inappropriate. A more suitable time might be more productive.

NOTE: **For this survey, language difficulties are not included under code K. See codes 0 and I.**

Code I: Reading and writing difficulty

Mark this code only when respondents state that they cannot read or write English (or French).

Do not use this code in cases where the respondent reads or writes another language. In those cases, use Code O, Language difficulties.

Code O: Language Difficulties

You were unable to complete the Background Questionnaire because the selected respondent did not speak English or French and an interpreter was not available.

OR

The selected respondent was not able to attempt the tasks in the Booklet because he/she does not have sufficient knowledge of English or French.

Code Z: Failed Core Tasks

This code is to be used only on the Main Tasks Booklet if the respondent answered less than two out of the six items correctly on the Core Tasks Booklet.

NOTE: **If the Booklet is not administered due to language difficulties, use Code O.**

Code U: Unable to Trace

Use Code U if you are unable to locate the respondent or the household after all tracing attempts have been made.

CODE A to H - DISABILITY CODES

Code A: Interview prevented due to learning disability

Enter this code if the respondent cannot do the survey due to a learning disability. The most common type of learning disability is dyslexia, a type of impairment that affect a person's ability to read. Another type of learning disability is Attention Deficit Disorder. Persons reporting a learning disability should still be encouraged to attempt the Core and Main Tasks Booklets.

Code B: Interview prevented due to mental or emotional condition

This includes psychological disorders, dementia, autism, nervous breakdowns, severe depression, schizophrenia, etc.

Code C: Interview prevented due to mental retardation

The selected respondent is mentally slow. This can be caused by a birth defect, a congenital condition such as Down's Syndrome or accident or injury to the brain that results in brain damage.

Code D: Interview prevented due to hearing impairment

If you encounter a respondent who is hearing impaired you can ask him/her to complete the background questionnaire themselves and to read the instructions for the tasks booklets. Use this code only if the respondent refuses to complete the background questionnaire due to a hearing impairment.

Code E: Interview prevented due to blindness or visual impairment

It will be impossible for respondents who are totally blind to complete the Tasks Booklets. Use Code E for these cases. If the respondent refuses to participate in the survey due to blindness enter this code on the background questionnaire as well, not code R.

Code F: Interview prevented due to speech or language impairment

Includes conditions such as severe stuttering and aphasia (impairment of verbal communication caused by brain damage.

Code G: Interview prevented due to physical disability

Physical conditions that interfere with the ability to complete the Tasks Booklets include paralysis, amputation of hand/arm, lack of muscle control.

Code H: Interview prevented due to other disability

Includes Alzheimer's disease, senility and any other condition not covered in codes A to G above.

SECTION 7 - REVIEW EXERCISE

Circle the letter beside the correct answer or enter the correct response.

1. The respondent has a hearing impairment. You should:
   - (a) End the interview and mark Code R in the Final Status Boxes on the BQ.
   - (b) Request that he attempt the Core and Main Tasks Booklets only.
   - (c) Ask the respondent to read the Background Questionnaire and tasks booklets himself.
   - (d) b) and c) above.

2. Indicate the final status code to be used on the Background Questionnaire beside each description. Print in BLOCK letters.
   - (a) __ An interview is prevented due to sickness in the household.
   - (b) __ The selected respondent has moved and you are unable to trace him.
   - (c) __ The respondent answered all questions on the BQ with the exception of 3.
   - (d) __ The respondent is absent for the entire survey period.
   - (e) __ Although you have made many attempts to contact the respondent and know the telephone number is correct, you are not able to arrange an interview.
   - (f) __ The respondent is unwilling to complete the Background Questionnaire.
   - (g) __ A respondent who is blind refused to answer theBackground Questionnaire.

3. David completed the Background Questionnaire. There is evidence that he attempted all the questions in the Core Tasks Booklet but he answered only one correctly. What is the final status of the Core Tasks Booklet?
   - (a) Code P
   - (b) Code X
   - (c) Code R

4. How would you mark the final status for the Main Tasks Booklet in the case above?
   (a) Code R
   (b) Code Z
   (c) Code 0
   (d) Code P

5. Due to language difficulties, a respondent was unable to complete the Background Questionnaire. What is the final status code of the Questionnaire?
   (a) Code K
   (b) Code P
   (c) Code O

6. Due to language difficulties the respondent was unable to attempt any of the Core Tasks Booklet. What are the final status codes for the Core Tasks and the Main Tasks Booklets?
   (a) Codes R and R
   (b) Codes O and Z
   (c) Codes O and O

7. The respondent completes the Background Questionnaire, then attempts all items on the Core Tasks Booklet. He answers three of these questions correctly, but refuses to attempt any of the items on the Main Tasks Booklet. Record the final status codes on the lines below.

   Background Questionnaire ______________________
   Core Tasks Booklet ______________________
   Main Tasks Booklet ______________________

8. When you attempt to contact the respondent using the telephone number on the label, you find that the respondent has moved. You are unable to trace the respondent. Record the final status codes on the lines below.

   Background Questionnaire ______________________
   Core Tasks Booklet ______________________
   Main Tasks Booklet ______________________

9. You administer the Background Questionnaire to the respondent, but due to a physical disability she is not able to attempt the completion of the Core Tasks and the Main Tasks Booklets. Record the final status codes on the lines below.

   Background Questionnaire ______________________
   Core Tasks Booklet ______________________
   Main Tasks Booklet ______________________

8 PROGRESS AND COST REPORT

8.1 Progress and Cost Report - Interviewers

You will be required to keep a record of your daily progress, hours worked and other expenses throughout the survey period. These figures are to be entered on the Progress and Cost Report and reported to the Senior Interviewer on specific dates.

Instructions for completing the Progress and Cost Report are outlined below:

R.O./Units Assigned

Enter the name or number of your R.O. and the number of units in your assignments.

Interviewer's name and reporting time

Enter your name and the reporting time assigned to you by your Senior Interviewer. Make sure to report at the appointed time since the Senior Interviewer will be collecting reports from each interviewer under her supervision.

You will report **weekly totals** to your Senior Interviewer on:

- Tuesday, September 6, 1994
- Monday, September 12, 1994
- Monday, September 19, 1994
- Monday, October 3, 1994 (Final Report)

NOTE: **If you find that you have time to conduct interviews for the International Adult Literacy Survey during LFS week, include these interviews in the total for Monday October 26th and report them as part of the total for Report 4.**

Senior Interviewer's Name and Telephone Number

Enter the name and telephone number of the person to whom you will be reporting.

Final Status of Documents

For each of the survey dates listed, enter the number of units in the appropriate final status code columns. Final status means that no further action will be taken on that document. All codes must be reported. Please note that Code N should not be used before the last week of the survey.

Interviewer Costs

Interviewer costs are separated into hours worked, distance travelled (in kilometres) and any other allowable expenses (e.g., telephone, meals, parking, etc.). The information recorded in these columns must be the same as you report on the F-85.

The first three rows of the report are reserved for training. Enter the hours spent on home study and classroom training (including travel time to and from the training session, distance travelled and any other allowable training expense).

The remaining rows are reserved for data collection. Enter your costs for each of the survey dates.

The "Report" rows represent the sum total for each of the Final Status Codes for that particular week.

An example of a completed Progress and Cost Report follows.

8.2 Interviewer Worksheet

The Interviewer Worksheet is printed on the back cover of your Interviewer's Manual. This worksheet should be used to keep a record of the hours you worked the distance travelled and other expenses for this survey. This information will help you complete your Progress and Cost Report and your F87.

8.3 Example of Completed Progress and Cost Report

9 TASK CODES

9.1 Task Codes

The project code for the International Adult Literacy Survey is **8176-0.**

Interviewer Task Codes

03 -Training (self-study and classroom training)

09 -Personal interviewing

10 -Telephone interviewing (includes calling for appointments and tracing)

17 -Assignment preparation

18 -Editing

10 TRANSMITTAL OF DOCUMENTS

10.1 Transmittal of Documents

Before packaging the survey documents for transmittal to your Senior Interviewer, ensure that:

- you have transcribed the R.O. Number and Docket Number accurately onto the Tasks Booklets and the Newspaper.
- the Task Booklets and the Newspaper are included in the respondent envelope along with the Background Questionnaire.

NOTE: **This also applies to non-response situations. ALL the respondent's documents, whether they were used or not should be returned to H.O. If for any reason you must use a non-response document for another respondent, please put a note in the respondent's envelope explaining this.**

It is your responsibility to ensure that completed questionnaires are packaged in accordance with shipping procedures outlined in the Security Manual. An excerpt follows:

> **CONFIDENTIAL MATTERS SHALL BE ENCLOSED IN TWO ENVELOPES OR COVERS WHEN BEING PREPARED FOR POSTAL TRANSMISSION OR BEING FORWARDED BY OTHER MEANS OUTSIDE A BUILDING.**
>
> The mode of shipment will be designated by your Regional Office.

A Transmittal Form must be completed for each shipment. It will serve as a packing slip. The procedures for the completion of the Transmittal Form appears at the back of the form.

Include all unused documents (Booklets, Newspaper) in the envelope. These documents must be kept together when transmitted to your supervisor.

You are to ship the documents for completed interviews at the end of each week of data collection except for the **first five** completed documents which should be sent to your supervisor as soon as possible so he/she can provide you with some feedback on your work. Your supervisor will provide feedback on any problems you may have completing the Background Questionnaire and/or the Tasks Booklets.

A completed example of the Transmittal Form appears on the next page.

10.2 Example of Completed Transmittal Form

Once the survey is completed, dispose of unused survey materials as follows:

SHIP TO R.O. (Attention: Project Supervisor, Int. Adult Literacy Survey	DESTROY:
- blank Background Questionnaires	- training materials
- blank Core and Main Tasks Booklets	- Interviewer's Manual
- blank Newspapers	- Tasks Administration Guide
	- Code Sheet
	- Edit Summary Sheet

10.3 Transfer procedures

Transfer procedures will be necessary if you are successful in tracing a respondent who has moved outside of your assignment area.

If you have the new address of the respondent telephone your Senior Interviewer to explain the situation. The Senior Interviewer will decide if you are still in the best location to conduct the interview with the respondent or if the case should be transferred to another interviewer.

If you are unable to trace a respondent, enter Code U on all documents and consider the case ended.

A. TRANSFER "OUT" PROCEDURES

If your Senior instructs you to transfer a case proceed as follows:

1. Enter all available information (new address, phone number) regarding the respondent in the Record of Calls section on the Cover page of the Background Questionnaire.
2. Indicate that this is a transfer case on the respondent's envelope and on the BQ.
3. Draw a line through the identification numbers for this unit on the Assignment Control List and write "transferred out" in the area for comments.
4. Indicate that a unit was "transferred out" of your assignment on the Progress and Cost Report.

B. TRANSFER "IN" PROCEDURES

The Interviewer who receives the unit will:

1. Manually enter the identification numbers for this unit onto his/her Assignment Control List and write "transfer received" in the area for comments.
2. Attempt to interview the selected respondent. **(The Interviewer must record the final status codes in Item 5 of the Background Questionnaire and include this document in her Progress Report to her Senior Interviewer)**
3. Indicate that a unit was "transferred into" his/her assignment on the Progress and Cost Report.
4. Ship the completed documents to your Senior Interviewer with your regular shipment of work.

11 QUESTIONS AND ANSWERS

11.1 Introduction

The most effective method of achieving our objectives is to obtain and maintain the support of the Canadian public. Collecting information is our basic function and we must become increasingly attuned to public relations as more demands are placed on us and, in turn, on the public many of whom are becoming more concerned about the invasion of their privacy.

Generally, if a respondent is treated well, he/she will respond well. Good respondent relations go beyond courtesy. It is important to realize that a respondent should know and indeed, has a right to know, certain information pertaining to the survey.

The following questions and answers are designed to provide you with consistent and accurate answers to some of the questions commonly posed by respondents. Become familiar with all of the answers, so you can reply comfortably to similar questions asked while conducting the interview.

11.2 How did you select me for this survey?

A sample of people of all ages has been chosen in each province to give a good representation of the population.

11.3 Why me instead of someone else in the family?

You were selected to represent other Canadians with similar characteristics. For example, age, sex, and community size. I am not allowed to substitute one household member for another. So, if we cannot interview you, the people you represent in the sample will not be reflected in the survey results. Your answers will represent the answers of over a thousand other Canadians.

11.4 Am I obligated to take part in the survey?

Your voluntary participation is very important because you have been selected to represent people having similar characteristics. To ensure that all such people are represented in Canada, it is necessary that we interview all respondents who have been selected.

11.5 Can we do the interview over the telephone?

No, part of the survey will involve you looking at certain types of material. Therefore it is necessary for us to interview you in person.

11.6 I'm too old. Why don't you talk to my son; he could better answer your questions?

We need to interview men and women from all age groups. Your answers are important to us and we appreciate your participation.

11.7 How much time will the interview take?

It could take anywhere from 20 minutes to one and one-half hour to complete the interview.

11.8 Who is involved in conducting this survey?

The survey is being conducted by Statistics Canada on behalf of the National Literacy Secretariat and Human Resources and Development.

11.9 How many people are included in the survey?

Approximately 6,500 people from across Canada will be interviewed for this Survey. Your participation is very important.

11.10 Do you need this information? Couldn't you get it from another source?

Information available from other sources is very limited. This survey will provide more complete information on the literacy abilities and needs of the Canadian population that is necessary for program planning.

11.11 The Government has too much information already.

The data that is available through other sources is outdated or incomplete. Canadian society is dynamic and there are constant changes in our way of life, attitudes, social and economic conditions. It is important to monitor these changes so that the government can act accordingly.

11.12 Who uses the information?

Government and non-government agencies, universities, institutes and the educational system will use the information from this survey to identify how literacy affects our lives, such as - Are Canadians being prepared through our educational systems to function effectively? Are we limited by our literacy skills in becoming more productive or in pursuing other job opportunities?

11.13 What is the International Adult Literacy Survey?

The objective of the survey is to profile literacy levels among the Canadian population. Information from this survey will be used by government departments and other agencies to help plan programs suited to the needs of Canadians.

11.14 These surveys are a waste of taxpayer's money.

The government uses social statistics to help them implement social programs and policies. Therefore, it is important that the government be well informed so that they can use your tax money in the most effective way.

11.15 How will I know if all this will be kept confidential?

The Statistics Act guarantees that any information you may give us remains strictly confidential. All employees of Statistics Canada are under oath and are subject to severe penalties if they do not safeguard that confidentiality.

Statistics Canada goes to extraordinary lengths to protect the confidentiality of the information it gathers. Once completed, the questionnaires are considered confidential and no unauthorized persons outside of Statistics Canada are allowed to see them.

Stringent precautions are taken in transporting and processing the survey information to ensure that your answers are kept confidential. Once tabulated, survey results are published in the form of graphs or tables.

11.16 Does Statistics Canada release information about individuals to other agencies?

Statistics Canada employees are subject to severe penalties should they release information about particular individuals or businesses. Statistics Canada releases only statistical summaries through which the identity of individuals can never be revealed. All publications are carefully reviewed from this point of view. We are proud of the fact that throughout the history of Statistics Canada we have always succeeded in safeguarding the confidentiality of information we have obtained. Not even other government departments can have access to the personal information.

11.17 Can I get a copy of my Task Booklets?

The scoring of these Booklets is done in Head Office. I cannot give you a copy of your Tasks Booklet. The information in the Booklets must remain confidential if we are to accurately profile literacy. (If respondents insist on a copy, they must contact Statistics Canada to request a copy).

11.18 When will the results be published?

The results are expected to be published in 1995.

11.19 How will the results be given to the general public?

The results will be published in 1995 and will be available through our Statistics Canada Regional Offices. Information will be made available to researchers, educators and the media.

11.20 I thought that a literacy assessment had already been done? Why are you doing another one?

You are correct, we conducted a literacy survey in the Fall of 1989. In order to implement social programs and policies, information needs to be updated. Also, this survey is conducted internationally whereas the 1989 survey was conducted in Canada only.

11.21 What do you mean by literacy?

For the purpose of this survey, literacy is being profiled in the two official languages and has been defined as: "The reading, writing and numeracy skills necessary to use the printed material commonly encountered at work, at home and in the community."

11.22 Why are people paying so much attention to adult literacy?

There is a growing concern about the literacy skills of the Canadian population. This affects our ability to adapt to the changing demands of the labour market and to fully participate in Canadian society.

11.23 Are most adults with literacy problems immigrants?

No. There is no indication that most adults with literacy problems are immigrants. In investigating how literacy abilities are distributed within the Canadian population, such demographic variables as age, sex, ethnic origin, mother tongue, and income will be considered.

11.24 How can Canadians have literacy problems when we have compulsory schooling in Canada?

There are many reasons why people may not be fully literate. Some live in isolated areas and are not able to attend school, while others have to stay home to help their families. Still others need special assistance to overcome learning difficulties or social problems and this assistance is not always available or the problem not recognized. Also, it is possible that skills once known are lost through lack of use.

11.25 Didn't the federal government recently give a lot of money to literacy programs? What has happened with that?

The allocation of the money is for a 5-year plan. This survey will help to identify the types of programs that will best meet Canadian needs.

11.26 What programs are available to improve my skills (Who can I talk to about literacy programs)?

We have contacted provincial agencies and educational systems which provide such services and a copy of those programs within your area is available by calling (___) ____ - ______. (Refer to list of offices provided by your Regional Office.

APPENDICES

APPENDIX A: FINAL STATUS CODE SHEET
APPENDIX B: EDIT SUMMARY SHEET

INTERNATIONAL ADULT LITERACY SURVEY
FINAL STATUS CODE SHEET

CODE	TITLE	DEFINITION
X	Completed	• BQ - Contains no more than 5 unaanswered questions. • **Core & Main** - There is evidence that the respondent attempted **each** question. The response need not be correct.
P	Partial	• **BQ** - More than 5 questions unanswered but most applicable question were answered. • **Core and Main Tasks Booklets** - There is no mark in the booklet for 1 or more questions.
R	Refusal	• **BQ** - Respondent refuses to participate or begins but does not provide enough information for the document to qualify as a partial. • **Core and Main** - Respondent refused: all questions are blank.
T	Temporarily Absent	• Respondent absent during the entire survey period.
N	No one home	• After numerous attempts telephone and personal attempts.
K	Interview prevented	• Due to death, short term illness or other unusual circumstances related to the household or the selected respondent. (EXCLUDE LANGUAGE DIFFICULTIES)
U	Unable to Trace	• Unable to trace respondent.
O	Language Difficulties	• Selected respondent or interpreter gives no information or provides less than is needed for a partial.
I	Reading & Writing Dif.	• Respondent states he cannot read or write English or French. If respondent reads or writes another language use code "O".
Z	Failed Core Tasks	• Used only on the Main Tasks Booklet when the respondent did NOT get 2 correct answers in the Core Tasks Booklet.

DISABILITY CODES

CODE	TITLE	DEFINITION
	Interview prevented due to:	
A	Learning Disability	• Examples of learning disabilities are dyslexia and Attention Deficit Disorder.
B	Mental/emotional condition	• Include dementia, autism, nervous breakdowns, depression, schizophrenia.
C	Mental retardation	• For example Down's Syndrome or caused by injury to the brain.
D	Hearing impairement	• Use only if hearing impaired respondent refuses to read and complete BQ and Core and Main Tasks Booklets.
E	Blindness	• Use for Tasks Booklets and for BQ if the respondent refuses to complete BQ due to blindness, visual impairement.
F	Speech impairement	• For example, stuttering or aphasia (impairement caused by brain damage).
G	Physical disability	• That interferes with the ability to complete booklets. For example, amputation of hand/arm, lack of muscle control.
H	Other disabilities	• Includes Alzheimer's disease, senility **and any other condition not covered in codes A to G above.**

REGIONAL OFFICE USE ONLY

Y	Interview Cancelled	Due to a lack of Interviewers, selected respondent cannot be interviewed.

January

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

February

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28					

March

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

April

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

May

S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

June

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

July

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

August

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

September

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

October

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

November

S	M	T	W	T	F	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

December

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

7-5030-427

Statistics Canada Statistique Canada

Canada

BEST COPY AVAILABLE

INTERNATIONAL ADULT LITERACY SURVEY
EDIT SUMMARY SHEET

A. EDITS TO BE DONE WHILE RESPONDENT COMPLETES MAIN TASKS

Interviewer Item 4 - Assistance provided - There must be an entry, "Yes" or "No" for this question.

Interviewer Items 6 and 7 - Start and end times BQ - These Interviewer Items are on pages 3 and 22 of the BQ. Ensure they are completed.

Flows (Go to...) - Flows throughout the questionnaire have been correctly followed. Circle skips.

Questions B2 to B5 - If B1 is a language other than English there must be an entry in questions B2 to B5. Ensure that the proper language code(s) have been entered in boxes in questions B2 to B5.

Questions B13 to B17 - If B14 to B17 completed ensure thare is an entry in B13.

Other - specify - Review all questions that have an entry in the "Other - specify" category and ensure that the entry you have made does not actually belong in one of the listed categories.

General edits - Ensure that:
- all entries are complete, correct and legible;
- an "X" has been marked in the appropriate circle;
- the time entries are recorded using the 24-hour clock; and

B. FINAL ENTRIES - TO BE DONE AFTER MAIN TASKS COMPLETED

Interviewer Item 8 - Ensure the language selected for the completion of the Tasks Booklets has been marked.

Interviewer Items 9 and 10 - Start and End Time Task Booklets - Start and end times for the Tasks Booklets have been entered in these Interviewer items on page 23 of the BQ.

Interviewer Item 2 - Total Number of Calls - Total number of calls from the Record of Calls and Appointments is recorded.

Interviewer Item 5 - Final Status - Ensure that a Final Status code has been recorded for all 3 survey documents: the Background Questionnaire and the Core and Main Task Booklets.

Transcribe these codes into the pre-printed Final Status area on respondent's envelope.

Interviewer Item 3 - Interviewer's Name - Sign your name after editing and recording final status.

Non-interview section -If you have not entered Code X for all the documents in Item 5 - Final Status, explain the reason for the non-interview in the Non-Interview section on the last page of the BQ.

AFTER THE INTERVIEW, ENSURE THAT ALL THE DOCUMENTS ARE RETURNED TO THE RESPONDENT'S ENVELOPE WHETHER THEY WERE COMPLETED OR NOT.

75030-6010

Statistics Canada Statistique Canada

BEST COPY AVAILABLE

Canadä

ERRATA

INTERNATIONAL ADULT LITERACY SURVEY

As discussed at Head Office training for the International Adult Literacy Survey, it has proven necessary to make changes to the definition of Status Codes **P** and **R** as they apply to the Main Tasks Booklet ONLY. The definitions should now read as follows:

Main Tasks Booklet P - There is no mark in the booklet for 1 or more questions.

Main Tasks Booklet R - All questions have been left blank due to a refusal.

Please ensure that the following changes are made to survet documents.

FINAL STATUS CODE SHEET:

P	should read	Core & Main - There is no mark in the booklet for 1 or more questions
R	(under the definition of R for the background questionnaire, add	
		Core & Main - All questions have been left blank due to a refusal.

Also please add the following text in regards to code **Z** to clarify that this code is only applicable to the Main Task Booklet).

Z	should read	Used only on the Main Task Booklet when the respondent did NOT get 2 correct answers in the Core Task Booklet.

INTERVIEWERS MANUAL:

page 4 - 8 Question B13 -	Interviewer should probe fr the inclusion of the language of interview. Due to the wording of this question, the respondent may feel that it is understood and so a respondent who speaks both English and Italian might answer only "Italian" to this question, the interviewer should repeat the chosen categories to the respondent to confirm that English is in fact spoken (e.g. "So you speak English and Italian?")
Page 7 - 1 Code P - Partial Questionnaire Should Read:	
Main Task Booklet -	There is no mark in the booklet for at least 1 or more questions (is all questions are blank, another status code must be used).
Page 7 - 2 Code R: Refusal Add:	
Core & Main -	The documents receiving this code should be entirely blank. The reason must clearly be a refusal by the respondent.
Page 10 - 4	Regarding disposal of unused supplies. Please return Tasks Administration Guide to the R.O. if they are still in good condition. They may be useful to Head Office.
Page 11 - 5 Question 11.26	insert the number for the ABC Hotline 1-800-465-7323
Appendix A:	Refer to changes describved above in regards to the Final Status Code Sheet.

PROCEDURES MANUAL:

p.25 (sect. 7): delete paragraph 2 in section 7.3 part A

Respondents who wish to obtain information on literacy programs should be referred to the ABC Hotline, 1-800-465-7323.

p.31 -	Delete "NOTE TO PROJECT OFFICER..."
p.32 -	Regarding Code P "MAIN TASKS BOOKLET - There is a mark in the booklet next to at least one of the questions." This instructions is correct but to be consistent the following change may be advisable.
Should read:	Main Task Booklet - There is no mark in the booklet for 1 or more questions.
Code R: Refusal	add Core & Main - All questions have been left blank due to a refusal.
P.41 -	Send an E-mail message to Ron Dubeau to advise him of the shipment and their contents as they are sent.

TRAINING GUIDE:

page 33 -	Final Status should be XXZ
page 38 -	B13 should be 01 and 03 (English and Italian)
page 83 -	H4 should be 06 (never). Not applicable (08) is only relevant if reading is truly not applicable (in the case of blindness or some other disability with precludes the child from reading *other than lack of reading skills)).
Page 97 -	Sept. 7 Interview 1, status should be XXP

Appendix E:

Sampling Guidelines

Georges Lemaitre

BEST CO LE

Table of Contents

Introduction 5

1. General notions 5

2. Drawing a sample 6
   - 2.1 List sampling 6
   - 2.2 Area sampling 7
   - 2.3 Random-digit dialing 7
   - 2.4 Quota sampling 8

3. Stratification, sample size and sample allocation 8

4. Over-sampling 9

5. Telephone coverage of the general population 10

6. Nonresponse and refusals 10

7. Summary and recommendations 11

Introduction

The legitimacy of results from surveys of the population is based, first and foremost, on the application of statistical sampling theory. The ability to make statements about the general population based on information about a relatively small subgroup of that population underlies much of modern research in economics and social science and indeed, has become essential in developing and informing much of government policy.

In this respect, the International Survey of Literacy follows a well-established tradition. The value-added afforded by the possibility of carrying out international comparisons, however, requires that one pay special attention to aspects of sampling, to ensure that this value-added is realized. The following report will begin by introducing some basic notions of sampling before going on to describe various sampling methods and their advantages and disadvantages. The report concludes with some general recommendations about sampling methods for use in the literacy survey.

The aim in this is not to impose or suggest a common sampling methodology for all countries. Indeed, this would be impossible to achieve, because private survey organizations, who will be carrying out both the sampling and the survey in many countries, do not necessarily have the same means at their disposal as national statistical agencies. Rather it is to ensure that reasonably representative samples can be drawn in all countries, given constraints on resources and methods, in order to ensure that reliable national estimates can be produced and to avoid the possibility that questions might be raised about the international comparability of the results because of methods of sampling that are not commonly accepted.

In what follows, we will often be talking for convenience about sampling households and about various methods of reaching persons to be surveyed, which will include, among others, telephoning. It is well to keep in mind throughout the discussion that the sampled units for the Survey of Literacy will be, not households, but individuals, and that regardless of how persons are initially contacted, the literacy assessment will be administered by means of an on-site interview.

1. General notions

For the literacy survey, a representative sample of the entire population is required. Normally, this is achieved by means of what is known as a probabilistic sample.

A probabilistic sample is one in which every person in the target population has a chance of being selected into the sample and where the chance of being selected is known (for example, 1 in 50). It does not necessarily require that every person in the target population have the same probability of being selected into the sample, but simply that each person have at least a chance of being included in the sample. A person's chance of being selected into the sample is known as the probability of selection, or the sampling rate.

The reciprocal of a sampled personals probability of selection (50, in the example above) is known as the sample weight and is generally interpreted as the number of persons in the target population represented by the sampled individual. Thus, the fact that two persons in the sample have different probabilities of selection does not mean that the sample is not representative, but simply that each of them represents different numbers of persons in the target population.

To generate estimates of the number of persons in the target population having a particular characteristic, for example the number of persons with university degrees, one sums the weights of persons in the sample with university degrees. Likewise, the proportion of persons in the target population that are employed is estimated by summing the weights of persons in the sample who are employed and dividing by the sum of the weights of all persons in the sample.

The sample weights are thus, the means that allow one to make statements about the target population on the basis of just a sample of that population.

One essential point to retain about sampling is that, all things being equal, the reliability of an estimate from a sample depends almost entirely on the sample size. In other words, a sample of 1,000 persons will produce estimates for a population of 1,000,000 persons that are as reliable as a sample of the same size from a population of 100,000 persons. We will return to this point later in the report, when we examine the possibility of producing regional estimates (which may be a concern in some countries).

2. Drawing a sample

2.1 List sampling

In order to draw a sample one has to have a specific means of selecting persons, householdsl etc. into the sample. The simplest way of doing this is to have a complete list of all the units in the target population and to select every mth unit, where m depends on how large the sample is to be. If the population is of size N and the sample is to be of size n, then m=N/n. "m" is then the sample weight and "1/m" is the probability of selection.

In certain countries (e.g. Denmark), there exists population registers containing the names and addresses of all persons in the population, as well as, additional auxiliary information, such as age, sex, education, etc. Where such registers exist, drawing representative samples of the entire population or of population subgroups is a particularly simple task.

A second example of a list commonly used for sampling is the telephone directory. This list is generally not complete, because of unlisted numbers and households without telephones. It may also include some households more than once (i.e., households with more than one telephone number). Thus, some care needs to be taken in using this list, and procedures developed to offset its deficiencies.

2.2 Area sampling

In practice, a complete list of all the persons in the target population may not exist, so that other means have to be used. National statistical agencies often draw what are known as "area samples", to compensate for the absence of a complete list.

Area sampling is carried out by first drawing samples of small areas throughout the country. These are generally census enumeration areas whose geographic limits have been precisely identified in the previous census, with perhaps 50-250 households in them. The sample can then consist of either all the households in these selected areas, or a second stage of selection can be carried out in which only a certain fraction of households in these areas (say, 1 in 10) will be selected into the sample. In order to do this, a list of households in each selected area may be obtainable from municipal registers or can be compiled, perhaps by the interviewer, before the survey is actually conducted. The households to be interviewed would then be selected from this list. Alternatively, the interviewer can count off and interview each mth household in the selected area during the course of interviewing.

Thus, with area sampling, one effectively compiles a list only for selected areas, rather than for all households in the country. An additional stage of sampling might consist of randomly selecting one person from each household for interviewing.

The method of sampling described above, consisting of first selecting areas and then households to interview within each area, is commonly known as "cluster sampling" because the selected households tend to be geographically clustered. When interviews are conducted face-to-face, this method of sampling reduces travelling time and costs considerably. On the other hand, it generally does not yield estimates that are as reliable as methods that yield more dispersed samples. The reason is that persons who live in the same neighbourhood tend to "resemble" one another more than persons from different neighbourhoods. The effective sample size from such a sample is smaller because one is to a certain extent interviewing the same kind of persons more than once. The estimates produced from cluster sampling are still unbiased but one obtains less information from such a sample than from one that is more dispersed. Note, however, that if one selects only one household from each selected area, the sample will be widely dispersed and yield more reliable estimates, but at a greatly increased cost.

2.3 Random-digit dialing

A second method of sampling often used by private-sector survey organisations is random-digit dialing, in which telephone numbers are dialed at random, eligible and valid numbers identified, and persons or households at those numbers selected into the sample. With this method, unlisted numbers can be reached, but considerable time is wasted screening out non-existent or commercial numbers.

Note that with this method, although one is not specifically selecting every mth person on a list, the probability of selection can still be determined if one knows the total number of residential telephone numbers in the population, which is generally obtainable from the telephone company (a small adjustment needs to be made for households with more than one telephone). In addition, non-responding numbers and refusals remain part of the sample and are adjusted for by increasing the sample weights of persons or households who responded.

BEST COPY AVAILABLE

2.4 Quota sampling

Still another method of sampling in common use, especially for opinion and attitude surveys, is known as quota sampling. This method is used because it is easy to implement, cost-effective, and handles non-response and refusals in a particularly simple way. Under quota sampling, interviewers are assigned quotas of persons bearing certain specified characteristics (for example, age, sex, occupation) to interview. Often the quotas are chosen in such a way that the sample will contain the same proportion of persons with the given characteristics as the target population, although it is possible to assign quotas that are proportionally larger for certain groups than for others.

In quota sampling, interviewers contact households, generally by telephoning, either randomly or from a list, and screen into the sample persons of the specified characteristics until each quota is filled. Households which refuse or that are still non-responding after a certain number of contact attempts are dropped and additional households contacted until the sample contains the "correct" number of persons in each group.

Although each person in the general population (say, with a telephone) has a chance of being selected into the sample with this method, exact probabilities of selection cannot be assigned to a selected person or household, either a priori or a posteriori. Although quota sampling in its initial stages bears a close resemblance to the probabilistic sampling of telephone numbers, either from a list or by random dialing, it departs from probabilistic sampling once quotas begin to be filled or nonresponding households or refusals are encountered. Probabilities of selection in the later stages of quota sampling are impossible to calculate, because they are no longer just a function of predetermined sampling parameters but also depend on which quotas are filled and how many previously contacted households have nonresponded or refused. The exact sample design is thus, in a certain sense unknown, which makes it impossible to apply sampling theory to the estimates produced from such a sample.

In practice, sample weights can nonetheless be assigned in quota sampling; these are equal to the number of persons in a group present in the general population divided by the number of persons in the quota for that group. However, the reciprocal of the sample weight cannot actually be interpreted as a probability of selection.

Nonetheless, quota sampling can yield estimates that compare favourably with those from probabilistic sampling methods, especially when the sample sizes are relatively small (at most 2,500-3,000 units) and the characteristics defining the quotas are good predictors of the variable of interest. The estimates from quota sampling are in general biased, but the bias can be small.

3. Stratification, sample size and sample allocation

Often in sampling, one wishes to ensure that a set of predetermined groups is represented in the sample. For example, one may wish to ensure that the sample includes persons from each region of the country.

The easiest way to do this is to stratify the sample, that is, to divide up the country into a number of geographically distinct areas (strata) and to draw a sample independently from each stratum. The probability of selection can then be made to

BEST COPY AVAILABLE

vary from stratum and stratum, so as to ensure adequate rural representation, for example. Alternatively, quotas can be established for each stratum independently.

The strata described here need to be distinguished from the clusters discussed earlier. A stratum is an area that possesses its own sample and will, therefore, be directly represented in the sample. A cluster, on the other hand, is a sampled area: unlike strata, not all clusters are present in the sample. Households from a cluster will be present in the sample only if that cluster is selected into the sample. A stratum generally consists of a large number of clusters, only one or several of which will be included in the sample.

A stratum need not be made up of contiguous areas. For example, a stratum could be formed by grouping together a set of clusters known to contain a high proportion of residents of a particular social group. A sample of one or more clusters from this stratum would then ensure adequate representation of this group in the sample.

It is common for national samples to include some form of geographic stratification. Stratification to ensure adequate representation of other characteristics, such as persons with high or low education, for example, is more difficult, because such characteristics are not in general strongly geographically localized.

Often national surveys must produce reliable estimates not only at the national level, but at the regional level. For some countries, the total sample for the International Survey of Literacy may be large enough to support regional estimates.

As noted earlier, the reliability of an estimate, all things being equal, depends almost exclusively on the sample size. Thus, estimates of comparable reliability for all regions can be achieved by allocating the same sample size to all regions (equal allocation). On the other hand, the reliability of national estimates is maximized by allocating a sample to each region that is proportional to its population (proportional allocation).

Since reliable national estimates can only be achieved at the expense of the reliability of regional estimates (and vice-versa), surveys often compromise by allocating regional samples that are somewhere in between equal and proportional allocation. One such compromise, for example, would be square-root allocation, in which the sample allocated to a region is proportional to the square root of its population.

4. Over sampling

Surveys that are targeted at particular socio-economic groups or where it is of interest to obtain a over-representation of a particular group of interest sometimes resort to over-sampling of these groups, especially when it is known that normal proportional sampling will not yield samples for the group of sufficient size to be useful for analysis.

As noted earlier, an overrepresentation of a particular group in the sample does not mean that the sample is not representative of the general population. The representativity is maintained by assigning to persons in these groups a higher probability of selection (i.e. a smaller sample weight) than other persons in the sample.

Over-sampling is generally carried out by selecting a larger number of households than is actually required for the sample and then carrying out a second stage of sampling in which larger sampling rates are applied to households containing members of the group to be over-sampled.

BEST COPY AVAILABLE

5. Telephone coverage of the general population

Whatever sampling method is used to draw a sample, the question of whether or not the method used is capable of reaching every one in the target population arises. In other words, does every person in the population of interest have a chance of being selected into the sample?

For sampling by telephone, this is clearly not the case, because not all households have telephones. This may not matter if the proportion of households without telephones is relatively small or if the characteristic of interest is uncorrelated with the fact of having a telephone.

In developed countries, telephone coverage is generally fairly complete (over 97% in Canada, approximately 94% in France). However, the characteristics of the population without telephones tend to be substantially different from that of the population with telephones. For example, in Canada, persons in households without telephones tend to have a much higher unemployment rate than that of the general population. This suggests that the non-telephone universe may tend to be concentrated among persons at the low end of the literacy scale. Thus, the selection of the sample by means of telephone numbers may leave out a certain portion of the population that is of particular interest to the survey.

Although sampling of telephone numbers is a convenient and costeffective way of choosing the sample, the cost benefits need to be weighed against the certainty that a perhaps not insignificant portion of the low-literacy portion of the general population may be missing from the sample.

6. Non-response and refusals

In any survey, a certain proportion of households or persons called or visited by an interviewer are either absent at the time of the call or visit, or refuse to be interviewed. Ideally, one would like to keep this portion of the sample relatively small.

For probabilistic sampling, it is desirable to have some idea in advance of the proportion of households likely to non-respond or refuse, so that the sample selected is large enough to yield the number of completed interviews desired. This is not an issue for quota sampling, because with this method, sampling continues until all the quotas are filled.

With either method, the composition of non-responding households and refusals is a matter of particular interest. Here the ideal situation would be that such households or persons are randomly distributed in the general population, that is, in statistical parlance, are "ignorable". In the Canadian literacy survey, nonrespondents were not concentrated among any particular groups.

For probabilistic sampling, "ignorable" non-response would permit a simple non-response adjustment in which the sample weights of responding persons would be increased to compensate for the nonrespondents and refusals. In quota sampling, "ignorable" nonresponse would mean that replacing a non-responding unit or a refusal would not have an adverse effect on the estimates.

Problems would arise in either case if persons at a particular literacy or numeracy level tended to non-respond or refuse more than persons at other levels. Quota sampling can actually provide some protection against this, if the quotas are based, at least in part, on a characteristic that is strongly correlated with literacy, in this case, educational attainment or mother tongue. In other words, quota sampling can guarantee the presence in the sample of a fixed number of persons at each educational level or with a mother tongue other than the language of the country.

In probabilistic sampling, on the other hand, the number of persons at each educational level will depend on both sampling variability and on the nature of non-response and refusals. However, there exist estimation methods that can compensate largely for this uncertainty in probabilistic methods.

Most surveys collect information about demographic characteristics such as age and sex, for which there exists reliable external information. For example, there generally exist very reliable estimates of the population obtainable from demographic sources (census, birth and death registers, etc.). The usual practice in probabilistic surveys is to modify the sample weights so that the estimates of the population by sex and age group obtained from the sample exactly agree with those obtainable from external sources. In this way, the sample "inherits" the reliability of the external data on population. In addition, the improvement in reliability for population tends to be transmitted to characteristics other than population. These procedures generally go under the name of "poststratification" or "benchmarking".

For the literacy survey, given the strong association between literacy and educational attainment, it would be desirable to carry out this kind of adjustment for the sample estimates of persons at each educational attainment level (as well as, by age group and sex). Many countries in the literacy survey may possess counts or estimates of the number of persons at each level of educational attainment (obtainable, say, from labour force surveys) which can be used to ensure that sample estimates at each level obtained from the survey agree with estimates from other sources where nonresponse is much less of a problem and where sample sizes (and therefore, the reliability) of the estimates are much greater.

7. Summary and recommendations

Both probabilistic and quota sampling methods are likely to be available in most countries. National statistical agencies generally use probabilistic methods, whereas private surveying organisations often use quota sampling.

Quota sampling is simple, cost-effective and can guarantee adequate representation in the sample of all levels of educational attainment. In many conditions, it can yield estimates that are relatively reliable compared to probabilistic methods, especially if samples are not large and the quotas are based on characteristics that are strongly associated with the variable of interest. However, quota sampling is not based on systematic randomization of the population units and there is always the risk that estimates from quota sampling may be biased, although the bias may be small. Moreover, it is impossible to determine ahead of time what the extent of the bias might be. Probabilistic sampling methods, on the other hand, guarantee unbiased estimates, provided non-response or refusals are "ignorable".

There is an additional argument against quota sampling that needs to be considered. In a survey such as the literacy survey that will be used to make international comparisons and from which will flow analyses that may influence national policy, one cannot afford the risk that the survey results will be called into question because of doubts about the sampling methodology. For this reason alone, it is **strongly recommended that probilistic sampling methods be adopted for the literacy survey.**

There are costs associated with this choice, because probabilistic methods are generally more expensive. In addition, they require a closer supervision of interviewers to ensure that non-responding households and refusals are not replaced by other units. However, if there is no alternative to quota sampling, it is essential that the quotas be at least partly based on educational attainment (and native and non-native mother tongue, if a country contains a significant immigrant population).

Among probabilistic methods, the two likeliest options are area sampling and telephone sampling. Because it is the only method that guarantees full coverage of the general population, it is **recommended that area sampling be the method of choice for the literacy survey.**

In practice, this may not always be possible, because private surveying organisations may not always have the infrastructure in place to carry out this kind of sampling, and may have to develop the capability at a certain additional and perhaps prohibitive cost to the client. It may be of interest here that Westat, the private group which carried out the Young Adult Literacy Assessment in the United States for the Educational Testing Service, did use area sampling for this survey, so that area sampling by private survey companies is not unprecedented.

If telephone coverage is extensive, countries may choose to adopt telephone sampling, while recognizing that a certain portion of the low literacy population will likely be excluded from the sample by this method. In countries where telephone coverage is more limited, area sampling is the only method that will guarantee an adequately representative sample of the general population.

Area sampling generally means cluster sampling, and it is essential that some **care be taken to ensure that the sample is sufficiently dispersed.** Selecting few but large clusters reduces travel time and costs considerably, but generally has an adverse effect on sampling variability. On the other hand, selecting many small clusters results in a more dispersed sample and therefore more reliable estimates but at the price of greatly increased costs.

In telephone sampling, the sample tends to be "naturally" dispersed, so that selecting many small clusters in area sampling may have cost implications that are comparable to sampling telephone numbers. In practice, the actual size of sample clusters will be largely determined by the information available on geography, usually from census sources. The essential point to retain is that once clusters are selected, a simple and practical way to select households within the selected clusters must be developed.

Finally, **it would be advisable to "benchmark survey estimates of the population by sex, age group and educational attainment to external estimates or counts that are more reliable,** provided theseare available. Strictly speaking, this is more related to estimation than to sampling methodology. However, it is an important point to keep in mind, because "benchmarking" can considerably offset the disadvantages of clustered samples that are normally associated with area sampling, as well as ensure that nonresponse and refusals do not overly distort the representativity of the sample.

Appendix F:

International Data Structure

(NOTE: Due to the length of the record layout, only an excerpt follows. To obtain a complete copy, contact the Special Surveys Division at Statistics Canada.)

```
                                                                                 PAGE    1
FIELD  ACRONYM   LEN  POSITION   QUESTION AND VARIABLE DESCRIPTIONS         UNWTD/WEIGHTED

                                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
                                 X                                             X
                                 X                                             X
                                 X         COMMON SECTION BEGINS               X
                                 X                                             X
                                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

    1  CNTRID      2  0001-0002  Country Identification Code

                                 01 Canada English
                                 02 Canada French
                                 03 Switzerland German
                                 04 Switzerland French
                                 05 Germany
                                 06 USA
                                 07 Ireland
                                 08 Netherlands
                                 09 Poland
                                 10 Not available
                                 11 Sweden
                                 12 Australia
                                 13 New Zealand
                                 14 England
                                 15 Northern Ireland
                                 16 Belgium (Flanders)

                                 NOTE: This is a required field. It is forced
                                       to  the output  file  and  has  to  be
                                       generated  by  a   program   manually.
                                       Thus, there is no reason why it should
                                       be missing and consequently, no values
                                       for  "not stated"  or  "missing"  have
                                       been  included.  For most  countries,
                                       this is a  simple process,  for Canada
                                       Switzerland and Norway, extra care must
                                       be taken.  For instance,in Switzerland,
                                       in order  for  a record to have a code
                                       "04",    you   must    first    check the
                                       language  of  test  (TBLANG),  if this
                                       variable is  equal  to French (code 2)
                                       then CNTRID is 04.Conversely if TBLANG
                                       is German (code 3) then CNTRID must be
                                       03   (irrespective  of  the  value  in
                                       BQLANG).  Likewise,  if TBLANG was not
                                       collected    (because  the  respondent
                                       refused the   test   or  some other such
                                       circumstance) then the value in BQLANG
                                       will drive the value of CNTRID  (i.e.,
                                       if BQLANG = 2, then CNTRID = 04, if it
                                       is equal to 3, then CNTRID = 03). This
                                       means that the coding of this variable
                                       is dependent on  TBLANG and,  only  if
                                       TBLANG  is  not  available,  does  it
                                       depend on the  value  in  BQLANG.  For
                                       countries   other   than   Canada   or
                                       Switzerland,   this distinction is not
                                       relevant.

                                       IALS1 countries include: Canada, Switzerland, Germany,
                                       USA, Netherlands, Poland, and Sweden.
```

BEST COPY AVAILABLE

```
                                                                   PAGE    2UNWTD/WEIGHTED
FIELD  ACRONYM   LEN  POSITION   QUESTION AND VARIABLE DESCRIPTIONS
    2  BOOKID      2  0003-0004  Booklet Identification Number

                                    Blank - No Booklet Administered
                                 01 Booklet 1
                                 02 Booklet 2
                                 03 Booklet 3
                                 04 Booklet 4
                                 05 Booklet 5
                                 06 Booklet 6
                                 07 Booklet 7
                                 08 US Booklet 1
                                 09 US Booklet 2
                                 10 US Booklet 3
                                 11 US Booklet 4
                                 12 US Booklet 5
                                 13 US Booklet 6
                                 14 US Booklet 7
                                 15 US Booklet 8
                                 16 US Booklet 9
                                 17 US Booklet 10
                                 18 US Booklet 11
                                 19 US Booklet 12
                                 20 US Booklet 13
                                 21 US Booklet 14
                                 22 US Booklet 15
                                 23 US Booklet 16
                                 24 US Booklet 17
                                 25 US Booklet 18
                                 26 US Booklet 19
                                 27 US Booklet 20
                                 28 US Booklet 21

                                 NOTE: This  field  is  equivalent  to  the
                                       booklet  number  which  has  been
                                       administered  to the  respondent.  In
                                       order  to remove any  ambiguity,  with
                                       regards  to  the  IALS  booklets,  the
                                       following  block  compositions  were
                                       used:

                                          IALS Booklet# = IALS Block#

                                          1 = 1, 2, 4
                                          2 = 2, 3, 5
                                          3 = 3, 4, 6
                                          4 = 4, 5, 7
                                          5 = 5, 6, 1
                                          6 = 6, 7, 2
                                          7 = 7, 1, 3

                                          The US books were constructed using  a
                                          1  NALS  2  IALS  Block  structure.
                                          Therefore, the US booklet structure is
                                          markedly different:

                                          US  Booklet#  =  Block#  (I#  =  IALS
                                          Block#;  N# = NALS Block#; NALS blocks
                                          are  numbered 1 and  8-13  to  conform
                                          with  internal  ETS  numbering;  IALS
                                          Block  numbers conform to  those  used
                                          above).

                                           1 = N1, I1, I2
                                           2 = N8, I2, I3
                                           3 = N9, I3, I4
                                           4 = N10, I4, I5
                                           5 = N11, I5, I6
                                           6 = N12, I6, I7
                                           7 = N13, I7, I1
                                           8 = I1, N10, I3
                                           9 = I2, N11, I4
```

BEST COPY AVAILABLE

```
FIELD  ACRONYM   LEN  POSITION   QUESTION AND VARIABLE DESCRIPTIONS          PAGE     3UNWTD/WEIGHTED
    2  BOOKID      2  0003-0004  Booklet Identification Number                          (CONTINUED)

                                 NOTE:     10 = I3, N12, I5
                                           11 = I4, N13, I6
                                           12 = I5, N1, I7
                                           13 = I6, N8, I1
                                           14 = I7, N9, I2
                                           15 = I1, I4, N12
                                           16 = I2, I5, N13
                                           17 = I3, I6, N1
                                           18 = I4, I7, N8
                                           19 = I5, I1, N9
                                           20 = I6, I2, N10
                                           21 = I7, I3, N11
    3  IDNO        9  0005-0013  Record Identification Number

                                 NOTE: This    variable   can    contain   any
                                       combination      of      alpha-numeric
                                       characters.  The record identification
                                       number should UNIQUELY point to a
                                       specific  respondent (record)  and  no
                                       duplicates  should  exist   within   a
                                       country.  A further  concern  regards
                                       the  eventual   need   to   link   the
                                       International  Data file back to  each
                                       individual  country files.   With this
                                       point in mind, this variable should be
                                       the same on both the international and
                                       your own national file.  Furthermore,
                                       you  should  endeavour  to  sort   the
                                       records  in the  same  order  on  both
                                       files  in order  to  facilitate  these
                                       eventual linkages.
    4  GENDER      1    0014     Gender of Respondent

                                 1 Male
                                 2 Female
                                 9 Not Stated

                                 NOTE: Since   this   is   an   observational
                                       variable (i.e.  the respondent needn't
                                       overtly  be asked  their  gender),  no
                                       category is offered for "Refused".

                                       Applicable to all IALS1 countries.
    5  AGE         2  0015-0016  Age of Respondent

                                 08 DK/Refused
                                 09 Not Stated
                                 15:99 Years

                                 NOTE: The  age variable should be the age of
                                       the respondent as of (Survey Date)
                                       rounded to the nearest  year.
                                       Given  the differences  in  collecting
                                       this  variable, such fine tunning  may
                                       not always be required (or possible).

                                       Survey Date = October 1, 1994 for IALS1
                                                     October 1, 1995 for IALS2

                                       Applicable to all IALS1 countries.
```

BEST COPY AVAILABLE

```
                                                                  PAGE     4UNWTD/WEIGHTED
FIELD  ACRONYM   LEN  POSITION   QUESTION AND VARIABLE DESCRIPTIONS
   6  MARITALS   1     0017     Marital Status of Respondent

                                1 Single
                                2 Married (including Common-Law)
                                3 Widowed
                                4 Divorced, Separated, etc.
                                8 DK/Refused
                                9 Not Stated

                                NOTE: Applicable to all IALS1 countries.
   7  COMMSIZE   1     0018     Respondent's Community Size, Rural/Urban

                                1 Rural
                                2 Urban
                                9 Not Stated

                                NOTE: It  is difficult  to  assess  the
                                      classification  of community size  due
                                      to the differing national  definitions
                                      of  what each category means  and  the
                                      various  ways in which  this  variable
                                      was  collected. For these  reasons,  a
                                      simple rural/urban split is  requested
                                      for  international  comparisons.  The
                                      definitions  of these two concepts  is
                                      left to each individual country  (much
                                      like the definition of the  boundaries
                                      between  income  quintiles  discussed
                                      below). Since this is an observational
                                      variable (i.e.  the respondent needn't
                                      overtly be asked their Community Size)
                                      no category is offered for "Refused".

                                      Applicable to all IALS1 countries.
   8  BQLANG     1     0019     Language of Background questionnaire

                                1 English
                                2 French
                                3 German
                                4 Dutch
                                5 Polish
                                6 Swedish

                                NOTE: This  variable will be  automatic  for
                                      most  countries. Only Switzerland  and
                                      Canada  have a  clear  distinction  to
                                      make.  Neverthless,  the  appropriate
                                      value  code should  be  inserted  into
                                      this  position.  It  refers  to  the
                                      language     of     the     Background
                                      questionnaire  (which may differ  from
                                      the  language  of  the  Task  Booklets
                                      (Cognitive items).
   9  STRTIMEH   2  0020-0021   Start time of BQ interview - Hours

                                01:24  Start time hour
                                99 Not Stated

                                NOTE: This  is the starting time in hours of
                                      the  BACKGROUND  questionnaire.    The
                                      format   is   standard    twenty-four
                                      hour (military) time.
```

```
                                                                    PAGE     5
FIELD  ACRONYM    LEN  POSITION   QUESTION AND VARIABLE DESCRIPTIONS            UNWTD/WEIGHTED
   10  STIMEMBQ     2  0022-0023  Start time of BQ interview - Minutes

                                  00:60 Start time minute
                                  99 Not Stated

                                  NOTE: This  is the starting time in  minutes
                                        of the BACKGROUND questionnaire.
   11  A1           1     0024    Were you born in (country of interview)?

                                  1 Yes
                                  2 No
                                  8 DK/Refused
                                  9 Not Stated

                                  NOTE: Applicable to all IALS1 countries.
   12  A5           2  0025-0026  Before you first immigrated to country of
                                  interview, what was the highest level of
                                  schooling you had completed?

                                     Blank - Not Applicable
                                  00 Education preceding the first level
                                  01 Education at the first level
                                  02 Education at the second level, first stage
                                  03 Education at the second level, second stage
                                  05 Education at the third level, first stage
                                  06 Education at the third level, first stage
                                  07 Education at the third level, second stage
                                  09 Education not definable by level
                                  10 No Schooling
                                  98 DK/Refused
                                  99 Not Stated

                                  NOTE: 05 of the type that leads to an award not
                                           equivalent to a first university degree
                                        06 of the type that leads to a first
                                           university degree or equivalent
                                        07 of the type that leads to a postgraduate
                                           university degree or equivalent

                                           Codes  00 to 09 are the ISCED codes as
                                           they  appear  in  the  UNESCO  ISCED
                                           abridged edition document.  Codes 10,
                                           98 and 99 have been added per force. A
                                           blank  should  only  be  used  if  the
                                           respondent  was  never  asked  this
                                           question  due  to  the  flow  of  the
                                           questionnaire (said "YES" to question
                                           A1   or "Citizen by birth" to question
                                           A3)

                                        Applicable to all IALS1 countries.
   13  A7           2  0027-0028  During your lifetime, how many years of
                                  formal education have you completed,
                                  beginning with afirst level, first stage,
                                  first year and not counting repeated years
                                  at the same level?

                                  00:90 Total years of schooling
                                  98 DK/Refused
                                  99 Not Stated

                                  NOTE: Applicable to all IALS1 countries.
```

BEST COPY AVAILABLE

```
                                                                             PAGE     6UNWTD/WEIGHTED
FIELD  ACRONYM   LEN  POSITION   QUESTION AND VARIABLE DESCRIPTIONS
   14  A8          2  0029-0030  What is the highest level of schooling you
                                 have completed

                                    Blank - Not Applicable
                                 00 Education preceding the first level
                                 01 Education at the first level
                                 02 Education at the second level, first stage
                                 03 Education at the second level, second stage
                                 05 Education at the third level, first stage
                                 06 Education at the third level, first stage
                                 07 Education at the third level, second stage
                                 09 Education not definable by level
                                 10 No Schooling
                                 98 DK/Refused
                                 99 Not Stated

                                 NOTE: 05 of the type that leads to an award not
                                          equivalent to a first university degree
                                       06 of the type that leads to a first
                                          university degree or equivalent
                                       07 of the type that leads to a postgraduate
                                          university degree or equivalent

                                          See  not for A5.   Given the flow into
                                          this  question,  the  "no   schooling"
                                          category is not needed. A blank should
                                          only  be used if  the  respondent  was
                                          never  asked this question due to  the
                                          flow  of the questionnaire (said  "00"
                                          to question A7)

                                       Applicable to all IALS1 countries.
```

Appendix G:

Example of a Scoring Rubric

BLOCK 2

Questions 8 - 10.

Weather Report

Question 8.

1 Circles the area of Norway/Sweden where the pattern of slanted lines appears (\\\\\\), and no other area of the map.*[Writing or underlining any part of text "rain will continue to soak southwestern Norway" is not an acceptable answer]*

7 Any other response

8 Newspaper is missing

0 Question refused / not done

Correct Response: **1**

Question 9.

1 18 (degrees(°))

7 Any other response

8 Newspaper is missing AND lines are blank

0 Question refused / not done

Correct Response: **1**

Question 10.

1 Caracas

7 Any other response

8 Newspaper is missing AND lines are blank

0 Question refused / not done

Correct Response: **1**

BLOCK 3

Questions 10 - 12.

Bicycle Frames

Question 10.

1 Mentions at least TWO of the following:
  - 530 (mm) OR 53 (cm)
  - 560 (mm) OR 56 (cm)
  - 580 (mm) OR 58 (cm)

7 Any other response

0 Question refused / not done

Correct Response: **1**

Question 11.

1 Mentions at least TWO of the following:
  - Knee (slightly) bent
  - Sole or equivalent part of foot on pedal
  - Pedal in the lowest position

7 Any other response

0 Question refused / not done

Correct Response: **1**

Question 12.

1 Underlines OR circles the sentence OR a part of the sentence that contains "The manufacturer is not responsible for failure, injury, or damage caused by improper completion of assembly".

7 Any other response

0 Question refused / not done

Correct Response: **1**

BLOCK 4

Questions 1 - 3.

Pain Reliever Label

Question 1.

1 Seven OR 7
7 Any other response
0 Question refused / not done

Correct Response: **1**

Question 2.

1 Mentions at least THREE of the following:
- (Before giving medication to children with) chicken pox
- (Before giving medication to children with) influenza (Reyes syndrome)
- (During) lactation
- (During) pregnancy
- If symptoms persist
- (Accidental) overdose

7 Any other response
0 Question refused / not done

Correct Response: **1**

Question 3.

1 8 (tablets)
7 Any other response
0 Question refused / not done

Correct Response: **1**

Appendix H:

Example of a Coding Scheme

PREFACE

This International Adult Literacy Major Field of Study Guide is based entirely on the MFS codes found in the Statistics Canada document entitled "User's Guide to the 1986 Census Data on Major Field of Study" prepared by G. A. Mori, M. E. Lalonde, B. Burke and A. Baril of the Housing, Family and Social Statistics Division. One of the first things that the you will notice is that the manual is VERY detailed. Unfortunately, it is based (obviously) on the Canadian education system. This means a certain compromise. For instance, we cannot use all of the detail in the manual; we must, in fact, code to a much less discreet denominator. Therefore, all of the codes which originally appeared in this document have been replaced by a series of numbers from 001 to 112. In addition, given the focus of the IALS, a series of supplementary codes are also included (numbered 120-141, plus 997 and 998) at the end of the document. In all, 135 codes are available. These codes, by and large, correspond to the Major headings of the MFS. Details of each heading are broken down into sub-headings and finally into fine detail further in this document. Despite all these details, it is important to keep in mind that we will only be coding to the broader headings.

Clearly, this manual is not perfect in an International context - it has many readily identifiable flaws which we will have to work around. One of the major flaws lies in its heavy emphasis on University or College levels. We will be ignoring these distinctions for the most part. In essence, we are coding to a SUBJECT not a LEVEL. Only when some subject overlap occurs in the manual should "level" enter the coding criteria (this happens most noticeably in the computer related courses whereby some are professional formations and others are strictly clerical). In order to effectively code the Major Field of Study, it is often useful to keep three additional variables handy: 1) Highest level of education attained; 2) whether the course (program) was taken for personal interest or work related reasons; and, 3) the respondents occupation.

Here then, follows the altered coding schemes to the MFS manual which you should use to code the International Adult Literacy Survey.

(NOTE: Due to the detail, only an excerpt follows. To obtain a complete copy of the coding schemes, contact the Special Surveys Division at Statistics Canada.)

Schematic Overview of the Broad Generic Level and Minor Group Classification Level of Major Field of Study

IALS Major Field of Study Codes 001 - 998

EDUCATIONAL, RECREATIONAL AND COUNSELLING SERVICES (001-009)

001	Education - General
002	Elementary - Primary Education
003	Secondary Education (Basic)
004	Secondary Education (Specialized)
005	Special Education
006	Non-teaching Educational Fields
007	Physical Education, Health and Recreation
008	Counselling Services and Personal Development
009	Other Education

FINE AND APPLIED ARTS (010-016)

010	Fine Arts
011	Music
012	Other Performing Arts
013	Commercial and Promotional Arts
014	Graphic and Audio-visual Arts
015	Creative and Design Arts
016	Other Applied Arts

HUMANITIES AND RELATED FIELDS (017-026)

017	Classics, Classical and Dead Languages
018	History
019	Library and Records Science
020	Mass Media Studies
021	English Language and Literature
022	French Language and Literature
023	Other Languages and Literature
024	Philosophy
025	Religious Studies
026	Other Humanities and Related Fields

SOCIAL SCIENCES AND RELATED FIELDS (027-039)

027	Anthropology
028	Archeology
029	Area Studies (Non Languages or Literature)
030	Economics
031	Geography
032	Law and Jurisprudence
033	Man/Environment Studies
034	Political Science
035	Psychology
036	Sociology
037	Social Work and Social Services
038	War and Military Studies
039	Other Social Sciences and Related Fields

COMMERCE, MANAGEMENT AND BUSINESS ADMINISTRATION (040-045)

040	Business and Commerce
041	Financial Management
042	Industrial Management and Administration
043	Institutional Management and Administration
044	Marketing, Merchandising, Retailing and Sales
045	Secretarial Science - General Fields

AGRICULTURAL AND BIOLOGICAL SCIENCES/TECHNOLOGIES (046-056)

046	Agricultural Science
047	Agricultural Technology
048	Animal Science Technologies
049	Biochemistry
050	Biology
051	Biophysics
052	Botany
053	Household Science and Related Fields
054	Veterinary Medicine/Science
055	Zoology
056	Other Agricultural and Biological Sciences/Technologies

ENGINEERING AND APPLIED SCIENCES (057-070)

057	Architecture and Architectural Engineering
058	Aeronautical and Aerospace Engineering
059	Biological and Chemical Engineering

060	Civil Engineering
061	Design/Systems Engineering
062	Electrical/Electronic Engineering
063	Industrial Engineering
064	Mechanical Engineering
065	Mining, Metallurgical and Petroleum Engineering
066	Resources and Environmental Engineering
067	Engineering Science
068	Engineering, n.e.c.
069	Forestry
070	Landscape Architecture

ENGINEERING AND APPLIED SCIENCE TECHNOLOGIES AND TRADES (071-082)

071	Architectural Technology
072	Chemical Technology
073	Building Technologies
074	Data Processing and Computer Science Technologies
075	Electronic and Electrical Technologies
076	Environmental and Conservation Technologies
077	General and Civil Engineering Technologies
078	Industrial Engineering Technologies
079	Mechanical Engineering Technologies
080	Primary Industries/Resource Processing Technology
081	Transportation Technologies
082	Other Engineering/Applied Science Technologies, n.e.c.

HEALTH PROFESSIONS, SCIENCES AND TECHNOLOGIES (083-098)

083	Dentistry
084	Medicine - General
085	Medicine - Basic Medical Science
086	Medical Specializations (Non-surgical)
087	Paraclinical Sciences
088	Surgery and Surgical Specializations
089	Nursing
090	Nursing Assistance
091	Optometry
092	Pharmacy and Pharmaceutical Sciences
093	Public Health
094	Rehabilitation Medicine
095	Medical Laboratory and Diagnostic Technology

096	Medical Treatment Technologies
097	Medical Equipment and Prosthetics
098	Other Health Professions, Sciences and Technologies, n.e.c.

MATHEMATICS AND PHYSICAL SCIENCES (099-109)

099	Actuarial Science
100	Applied Mathematics
101	Chemistry
102	Geology and Related Fields
103	Mathematical Statistics
104	Mathematics
105	Metallurgy and Materials Science
106	Meteorology
107	Oceanography and Marine Sciences
108	Physics
109	General Science

ALL OTHER N.E.C. (110)

110	All Other n.e.c.

NO SPECIALIZATION (111)

111	No Specialization

NO POSTSECONDARY QUALIFICATION (112)

112	No Postsecondary Qualification

UPGRADING (120-130)

120	Upgrading - General
121	Basic Education (Grades 1-8)
122	General Educational Development
123	High School Subjects
124	Post Secondary Upgrading
125	Pre-Vocational Upgrading
126	Basic Training for Skill Development
127	Basic Job Readiness Training
128	Orientation
129	Career Alternatives
130	University Transfer

PERSONAL DEVELOPMENT (131-138)

131	Personal Development - General
132	Home and Family
133	Consumer/Financial

134	Coping Skills
135	Communications Skills
136	Religion and Morals
137	Public Affairs, Community/Current Events
138	Driver Instruction

RECREATIONAL ACTIVITY (139-141)

139	Sports and Outdoor Recreation
140	Physical Fitness
141	Games

OTHER RESPONSE CODES (997-998)

997	Don't Know
998	Not Applicable

Major Field of Study Classification and Description of Elements

EDUCATIONAL, RECREATIONAL AND COUNSELLING SERVICES (001-009)

Education - General (001)

* **Education - General**

General studies in the skills and methods of teaching. This classification is often used in cases where no specialization in teaching had been specified.

Education
Teacher's (College Certificate)

Elementary - Primary Education (002)

* Elementary School Teaching - General

General studies in the skills and methods of teaching at the elementary school level without indication of specialization in any particular field of elementary school teaching.

Childhood Education
Elementary Education
Primary Education

* Elementary School Teaching - Specialized

Specialized studies in the skills and methods of teaching at the elementary school level. Persons specialized in teaching a particular subject at the elementary level are classified in this category.

Reading (Elementary Education)
Other (e.g., Physical Education, Arts)

* Pre-school, Kindergarten and Early Childhood Education

Studies in the skills and methods of early childhood education teaching pre-school and kindergarten.

Early Childhood Education

Kindergarten Education

Nursery Aid

Pre-school Education

Secondary Education (Basic) (003)

* Secondary School Teaching

Studies in the skills and methods of teaching at the secondary (high) school level. This classification was often used in cases where no specialization in teaching at the secondary school level had been specified.

Education Secondary

High School Education

Secondary School Education

* English Language Training

Studies in the skills and methods of teaching English as a first language at the secondary (high) school level or higher. This classification is also used in cases where teaching English was indicated but no level of teaching was specified.

Education - English

English Education

* French Language Training

Studies in the skills and methods of teaching French as a first language at the secondary (high) school level or higher. This classification is also used in cases where teaching French was indicated but no level of teaching was specified.

French Education

* Other Language Training

Studies in the skills and methods of teaching languages other than French or English.

Education - Languages

Language Education

* Mathematics - Science Teaching

Studies in the skills and methods of teaching mathematics at the secondary (high) school level or higher. This classification is also used in cases where teaching mathematics was indicated but no level of teaching was specified.

Computer in Education

Computer Education

Education - Math

Education Biology

Education Mathematics

Education Science
Math and Science Curriculum
Math Education
Mathematics and Education
Mathematics Education
Mathematics Teaching
Science Education
Science Teaching
Secondary Science

* Social Studies Teaching

Studies in the skills and methods of teaching the social sciences, (including history, sociology, psychology and law).

Education - Geography
Education - Social Studies
Education Geography
Education History
Education Social Studies
Geography Teaching
History for Secondary Schools
History Teaching
Secondary Education Social Studies
Social Studies
Social Studies Education

* Secondary Basic Curriculum - Other

Studies in the skills and methods of teaching, at the secondary (high) school level, basic curriculum subjects not specified above.

College Preparation
College Preparatory
Curriculum
Curriculum and Instruction
Curriculum Design
Curriculum Development
Curriculum Education
Curriculum Studies
Education - Curriculum
Education Curriculum - Studies
Education in Curriculum Instruction
Education Curriculum and Instruction
Education Curriculum Development
Educational Curriculum
Intermediate Education

Secondary Education (Specialized) (004)

* Adult/Continuing Education

Studies in the skills and methods of teaching adults in continuing education programs.

- Academic Upgrading
- Adult Education
- Continuing Education
- Extension Education

* Art and Fine Art Education

Studies in the skills and methods of teaching arts and fine arts.

- Art Education
- Dance Teachers' Training
- Drama in Education
- Fine Art Education

* Commercial/Business Education

Studies in the skills and methods of teaching commercial and business administration skills.

- Business Education
- Business Education Secretarial
- Business Teacher Education
- Commercial Teacher Training
- Office Education
- Teachers Degree Accounting

* Family/Life Education

Studies in the skills and methods of teaching home economics, household and domestic skills.

- Family Life Education
- Home Economics Education
- Human Relations

* Industrial/Vocational Education

Studies in the skills and methods of teaching occupational and vocational skills.

- Education Technical Teachers Certificate
- Education Vocational Art
- Industrial Arts Teacher Education
- Industrial Education
- Industrial Teacher Education
- Vocational Education
- Vocational Teacher Training

* Music Education

Studies in the skills and methods of teaching the music curriculum.

- Education Music
- Music and Education
- Music Education
- Music Teacher Training
- Music Teaching
- Musical Education

* Religious Education

Studies in skills and methods of pastoral or religious teaching.

- Christian Education
- Religious Education

Appendix I:

Additional Documentation

Chapter 4

Table I.1: **Response rate by age group, education level, and employment status: United States**

	Education					
Age/Employment	Primary school or less	Lower secondary	Upper secondary	Some post-secondary	University degree	Total
Age 16-35						
Employed	52	56	52	55	60	**56**
Unemployed	63	54	59	—	—	**55**
Other	65	58	51	46	53	**55**
Total	**60**	**56**	**52**	**53**	**60**	**55**
Age 36-55						
Employed	77	71	53	65	69	**67**
Unemployed	—	—	59	—	—	**63**
Other	74	71	63	65	65	**68**
Total	**75**	**70**	**63**	**65**	**69**	**67**
Age 56-65						
Employed	80	64	72	67	82	**74**
Unemployed	—	—	—	—	—	**76**
Other	86	68	71	67	71	**72**
Total	**85**	**66**	**71**	**67**	**80**	**73**

— Both numerator and denominator are less than 20

Table I.2: Results based on base and final weights by age group: United States

Scale/ Age/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Prose					
Age 16-35					
1	20.53	20.78	-0.25	1.33	-0.19
2	25.87	26.76	-0.89	1.92	-0.46
3	34.97	34.86	0.11	1.81	0.06
4 and 5	18.63	17.61	1.02	1.93	0.53
Age 36-55					
1	19.48	19.70	-0.22	1.54	-0.14
2	22.71	23.63	-0.92	2.47	-0.37
3	29.96	30.30	-0.34	2.43	-0.14
4 and 5	27.85	26.37	1.48	1.87	0.79
Age 56-65					
1	23.48	23.40	0.08	2.72	0.03
2	29.29	30.29	-1.00	3.14	-0.32
3	31.69	31.52	0.17	2.96	0.06
4 and 5	15.53	14.78	0.75	2.65	0.28
Document					
Age 16-35					
1	22.56	22.65	-0.09	1.34	-0.07
2	25.17	25.98	-0.81	1.82	-0.45
3	32.64	32.59	0.05	2.13	0.02
4 and 5	19.64	18.78	0.86	1.82	0.47
Age 36-55					
1	22.91	23.23	-0.32	1.43	-0.22
2	22.57	23.47	-0.90	1.85	-0.49
3	32.12	32.02	0.10	1.71	0.06
4 and 5	22.40	21.28	1.12	1.83	0.61
Age 56-65					
1	28.53	28.52	0.01	2.38	0.004
2	31.98	33.12	-1.14	3.44	-0.33
3	26.40	26.13	0.27	2.90	0.10
4 and 5	13.09	12.23	0.86	2.17	0.40
Quantitative					
Age 16-35					
1	22.06	22.40	-0.34	1.55	-0.22
2	24.70	25.61	-0.91	1.76	-0.52
3	32.66	32.72	-0.06	2.30	-0.03
4 and 5	20.57	19.26	1.31	2.08	0.63
Age 36-55					
1	19.27	19.48	-0.21	1.45	0.14
2	22.62	23.61	-0.99	2.18	-0.46
3	29.30	29.54	-0.24	2.11	-0.11
4 and 5	28.81	27.37	1.44	1.87	0.77
Age 56-65					
1	21.53	21.47	0.06	2.14	0.03
2	28.87	29.90	-1.03	3.65	-0.28
3	32.26	32.25	0.01	3.24	0.003
4 and 5	17.35	16.38	0.97	3.22	0.30

Note: The standard errors take into account sampling and measurement errors.

Table I.3: Results based on base and final weights, by education level: United States

Scale/ Education/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Prose					
Primary school or less					
1	74.39	73.82	0.57	5.77	0.10
2	18.58	18.77	-0.19	5.45	-0.03
3	6.49	6.87	-0.38	3.16	-0.12
4 and 5	0.54	0.54	0	0.49	0
Lower secondary					
1	49.89	46.47	3.42	2.61	1.31
2	26.70	28.04	-1.34	2.54	-0.53
3	18.92	20.25	1.33	2.41	-0.55
4 and 5	4.49	5.24	-0.57	1.66	-0.45
Upper secondary					
1	20.76	21.15	-0.39	1.94	-0.20
2	36.18	36.03	0.15	2.57	0.06
3	32.99	32.84	0.15	3.02	0.05
4 and 5	10.07	9.98	0.09	1.84	0.05
Some post-secondary					
1	8.04	8.55	-0.51	1.57	-0.32
2	25.95	26.24	-0.29	2.54	-0.11
3	39.84	39.78	0.06	2.65	0.02
4 and 5	26.16	25.43	0.73	3.03	0.24
University degree					
1	4.30	4.36	-0.06	1.29	-0.05
2	10.37	10.51	-0.14	1.57	-0.09
3	36.57	36.79	-0.22	3.97	-0.06
4 and 5	48.76	48.34	0.42	3.34	0.13
Document					
Primary school or less					
1	79.62	78.63	0.99	5.14	0.19
2	16.89	17.61	-0.72	5.07	-0.14
3	3.50	3.76	-0.26	1.89	-0.14
4 and 5	0	0			
Lower secondary					
1	51.01	47.78	3.23	3.50	0.92
2	24.89	25.85	-0.96	2.72	-0.35
3	18.48	19.96	-1.48	3.20	-0.46
4 and 5	5.62	6.41	-0.79	2.51	-0.31
Upper secondary					
1	24.61	24.89	-0.28	1.96	-0.14
2	36.57	36.41	0.16	2.31	0.07
3	30.08	29.97	0.11	2.69	0.04
4 and 5	8.73	8.73	0	1.96	0
Some post-secondary					
1	11.37	11.83	-0.46	1.15	-0.40
2	25.92	26.22	-0.30	2.63	-0.11
3	37.89	37.71	0.18	2.03	0.09
4 and 5	24.83	24.25	0.58	2.53	0.23
University degree					
1	6.71	6.76	-0.05	1.84	-0.03
2	11.49	11.68	-0.19	1.58	-0.12
3	40.39	40.54	-0.15	2.91	0.05
4 and 5	41.40	41.03	0.37	3.04	0.12

Table I.3: Results based on base and final weights, by education level: United States (Concluded)

Scale/ Education/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Quantitative					
Primary or less					
1	74.17	73.91	0.26	4.75	0.05
2	16.28	15.86	0.42	3.73	0.11
3	9.55	10.23	-0.68	3.86	-0.18
4 and 5	0	0	0	0	0
Some secondary					
1	49.88	46.96	2.92	4.06	0.72
2	23.81	24.73	-0.92	3.61	-0.25
3	20.62	22.07	-1.45	4.09	-0.35
4 and 5	5.69	6.25	-0.56	2.55	-0.22
Complete secondary					
1	21.39	21.71	-0.32	1.55	-0.21
2	36.50	36.46	0.04	3.47	0.01
3	32.05	31.83	0.22	2.70	0.08
4 and 5	10.05	10.00	0.05	2.09	0.02
Some post-secondary					
1	8.12	8.51	-0.39	1.27	-0.31
2	26.13	26.55	-0.42	3.10	-0.14
3	37.90	37.71	0.19	2.58	0.07
4 and 5	27.85	27.22	0.63	2.69	0.23
University degree					
1	4.51	4.63	-0.12	1.05	-0.11
2	9.53	9.80	-0.27	1.33	-0.20
3	33.42	33.64	-0.22	3.29	-0.07
4 and 5	52.53	51.93	0.60	3.18	0.19

Note: The standard errors take into account sampling and measurement errors.

Table I.4: Results based on base and final weights, by employment status: United States

Scale/ Employment/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Prose					
Employed					
1	15.64	16.04	-0.40	1.37	-0.29
2	24.62	25.54	-0.92	1.88	-0.49
3	33.82	33.96	-0.14	2.07	-0.07
4 and 5	25.93	24.47	1.46	1.66	0.88
Unemployed					
1	36.17	35.93	0.24	7.74	0.03
2	26.58	26.76	-0.18	8.26	0.02
3	28.09	28.98	-0.89	8.53	-0.10
4 and 5	9.16	8.33	0.83	4.25	0.20
Other					
1	33.26	32.62	0.64	2.37	0.27
2	25.67	26.72	-1.05	2.42	-0.43
3	27.91	27.91	0	1.69	0
4 and 5	13.16	12.74	0.42	1.61	0.26

Table I.4: Results based on base and final weights, by employment status: United States (Concluded)

Scale/ Employment/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Document					
Employed					
1	18.77	19.07	-0.30	1.09	-0.28
2	23.70	24.69	-0.99	1.56	-0.63
3	33.69	33.66	0.03	1.09	0.03
4 and 5	23.84	22.59	1.25	1.71	0.73
Unemployed					
1	41.53	41.25	0.28	6.32	0.04
2	19.49	19.65	-0.16	4.96	-0.03
3	28.86	29.29	-0.43	8.40	-0.05
4 and 5	10.12	9.81	0.31	3.32	0.09
Other					
1	35.90	35.49	0.41	1.91	0.21
2	29.78	30.43	-0.65	3.50	-0.19
3	25.09	24.87	0.22	3.13	0.07
4 and 5	9.23	9.22	0.01	2.02	0.004
Quantitative					
Employed					
1	16.46	16.84	-0.38	1.01	-0.38
2	23.22	24.25	-1.03	1.71	-0.60
3	32.11	32.33	-0.22	1.34	-0.16
4 and 5	28.20	26.58	1.62	1.32	1.23
Unemployed					
1	34.15	33.85	0.30	5.78	0.17
2	28.37	28.64	-0.27	7.56	-0.04
3	21.49	22.11	-0.62	6.93	-0.09
4 and 5	15.99	15.40	0.59	5.60	0.11
Other					
1	32.04	31.72	0.32	2.45	0.13
2	27.32	28.12	-0.80	3.06	-0.26
3	29.30	29.32	0.07	1.93	0.04
4 and 5	11.34	10.94	0.40	2.40	0.17

Note: Standard errors take into account of both sampling and measurement errors.

Table I.5 Results based on base and final weights, by sex: United States

Scale/ Sex/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Prose					
Male					
1	21.84	22.33	-0.49	1.64	-0.29
2	26.89	27.92	-1.03	2.08	-0.50
3	29.84	29.72	0.12	1.67	0.07
4 and 5	21.42	20.03	1.39	1.71	0.81
Female					
1	19.26	19.21	0.05	1.39	0.04
2	23.48	24.05	-0.87	1.72	-0.50
3	34.37	34.66	-0.29	1.89	-0.15
4 and 5	23.19	22.08	1.11	1.67	0.66

Table I.5 Results based on base and final weights, by sex: United States (Concluded)

Scale/ Sex/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Document					
Male					
1	24.88	25.28	-0.40	1.42	-0.28
2	23.66	24.63	-0.97	1.97	-0.49
3	30.66	30.53	0.13	1.84	0.07
4 and 5	20.80	19.57	1.23	1.84	0.67
Female					
1	22.41	22.39	0.02	1.15	0.02
2	26.14	26.96	-0.82	2.03	-0.40
3	32.26	32.21	0.05	1.91	0.03
4 and 5	19.18	18.44	0.74	1.44	0.51
Quantitative					
Male					
1	20.61	21.05	-0.44	1.57	-0.28
2	21.14	22.04	-0.90	2.21	-0.41
3	29.38	29.70	-0.32	1.76	-0.18
4 and 5	28.87	27.20	1.67	1.50	1.11
Female					
1	20.83	20.89	-0.06	1.56	-0.04
2	27.24	28.21	-0.97	1.62	-0.60
3	32.60	32.58	0.02	0.95	0.02
4 and 5	19.33	18.321	1.01	1.71	0.59

Note: Standard errors take into account of both sampling and measurement errors.

Table I.6: Results based on base and final weights, by race: United States

Scale/ Race/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Prose					
White					
1	11.90	12.51	-0.61	1.30	-0.47
2	24.28	25.20	-0.92	1.99	-0.46
3	36.18	36.08	0.10	1.75	0.06
4 and 5	27.65	26.21	1.44	1.49	0.97
Black					
1	38.73	34.78	3.95	4.55	0.87
2	35.01	36.28	-1.27	4.27	-0.30
3	21.78	23.88	-2.10	2.72	-0.77
4 and 5	4.48	5.06	-0.58	1.28	-0.45
Hispanic					
1	62.43	60.26	2.17	3.73	0.58
2	14.93	16.00	-1.07	4.18	-0.26
3	18.50	19.40	-0.90	3.03	-0.30
4 and 5	4.15	4.34	-0.19	1.45	-0.13
American Indian					
—					
Asian					
1	34.27	35.30	-1.03	4.72	-0.21
2	27.44	27.28	0.16	7.38	0.02
3	21.34	20.85	0.49	4.42	0.11
4 and 5	16.94	16.56	0.38	8.47	0.04
Other					
—					

Table I.6: **Results based on base and final weights, by race: United States** (Concluded)

Scale/ Race/ Literacy level	Base weight (1)	Final weight (2)	Difference of (1) and (2) (3)	Standard error (4)	(3)/(4) (5)
Document					
White					
1	15.36	15.96	-0.60	1.15	-0.52
2	24.98	25.83	-0.85	1.95	-0.44
3	35.34	35.08	0.26	1.21	0.21
4 and 5	24.32	23.13	1.19	1.70	0.70
Black					
1	42.60	38.71	3.89	3.95	0.98
2	32.86	34.09	-1.23	3.92	-0.31
3	19.65	21.70	-2.05	3.87	-0.53
4 and 5	4.89	5.50	-0.61	2.00	-0.31
Hispanic					
1	60.46	58.32	2.14	2.85	0.75
2	17.55	18.58	-1.03	2.76	-0.37
3	15.28	16.19	-0.91	3.07	-0.30
4 and 5	6.71	6.91	-0.21	2.11	-0.09
American Indian					
—					
Asian					
1	36.60	37.63	-1.03	5.63	-0.18
2	20.83	20.96	-0.13	6.42	0.02
3	31.03	30.27	0.76	5.77	0.13
4 and 5	11.53	11.15	0.38	4.37	0.09
Other					
—					
Quantitative					
White					
1	12.30	12.92	0.62	0.93	-0.67
2	23.45	24.32	-0.87	1.87	-0.46
3	34.98	35.06	-0.08	1.31	-0.06
4 and 5	29.28	27.69	1.59	1.45	1.10
Black					
1	40.20	36.55	3.65	3.56	1.03
2	36.60	37.58	-0.98	3.20	-0.30
3	18.58	20.58	-2.00	3.10	-0.64
4 and 5	4.62	5.30	-0.68	1.16	-0.58
Hispanic					
1	60.44	58.24	2.20	2.31	0.95
2	18.32	19.52	-1.20	2.78	-0.43
3	15.39	16.23	-0.84	2.23	-0.37
4 and 5	5.85	6.00	-0.15	1.59	-0.09
American Indian					
—					
Asian					
1	31.41	32.33	-0.92	5.56	-0.17
2	20.32	20.41	-0.09	4.47	-0.02
3	29.06	28.58	0.48	4.85	0.10
4 and 5	19.21	18.68	0.53	5.22	0.10
Other					
—					

Note: The standard errors take into account sampling and measurement errors.

— Unweighted cell size less than 20.

Table I.7: **Estimates from probit model predicting response**

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Own home	-0.338	0.046	-7.404	0.000	-0.427	-0.248
Income	-0.005	0.006	-0.785	0.432	-0.017	0.007
Hours of work	0.007	0.002	3.206	0.001	0.003	0.011
Multiple jobs (dummy)	-0.173	0.109	-1.582	0.114	-0.387	0.041
Income missing (dummy)	0.312	0.093	3.359	0.001	0.130	0.493
Work hours vary (dummy)	0.237	0.125	1.890	0.059	-0.009	0.482
Northeast (dummy)	-0.122	0.061	-2.005	0.045	-0.242	-0.003
Midwest (dummy)	-0.096	0.057	-1.675	0.094	-0.208	0.016
South (dummy)	-0.012	0.054	-0.229	0.819	-0.117	0.093
Central city (dummy)	-0.044	0.043	-1.023	0.307	-0.129	0.041
Metro area (dummy)	-0.108	0.069	-1.574	0.115	-0.242	0.026
Married (dummy)	-0.022	0.060	-0.366	0.714	-0.140	0.096
Single (dummy)	0.114	0.069	1.650	0.099	-0.021	0.249
Disability indicator (dummy)	0.065	0.294	0.222	0.824	-0.511	0.641
Temporary disability (dummy)	-0.716	0.384	-1.867	0.062	-1.468	0.036
In School (dummy)	-0.074	0.103	-0.712	0.477	-0.276	0.129
Child Care limits work (dummy)	-0.016	0.089	-0.184	0.854	-0.191	0.158
White (dummy)	0.007	0.217	0.034	0.973	-0.418	0.432
Black (dummy)	0.002	0.220	0.009	0.993	-0.430	0.434
Hispanic (dummy)	-0.074	0.222	-0.332	0.740	-0.508	0.361
Asian (dummy)	-0.202	0.229	-0.880	0.379	-0.651	0.248
Employed (dummy)	-0.283	0.106	-2.657	0.008	-0.491	-0.074
Unemployed (dummy)	-0.054	0.104	-0.521	0.602	-0.259	0.150
Age squared (dummy)	0.000	0.000	-5.024	0.000	0.000	0.000
Highest Education - primary (d)	-0.193	0.087	-2.228	0.026	-0.362	-0.023
Highest Education - Lower Sec. (d)	-0.045	0.062	-0.718	0.473	-0.167	0.077
Highest Education - Upper Sec. (d)	0.186	0.056	3.288	0.001	0.075	0.296
Some postsecondary (d)	0.161	0.059	2.721	0.006	0.045	0.277
Constant	0.209	0.252	0.830	0.406	-0.284	0.702

Table I.8: **Analysis of variance of weights implied by probit analysis presented in Table A4.7**

	Number of obs=4,875		R-squared		0.2267
	Root MSE		0.880254 Adj R-squared		0.2228
Source	Partial SS	df	MS	F	Prob > F
Model	1,101.442	24.000	45.893	59.230	0.000
Race*education	1,101.442	24.000	45.893	59.230	0.000
Residual	3,758.013	4,850.000	0.775		
Total	**4,859.454**	**4,874.000**	**0.997**		

Appendix J:

Additional Documentation

Chapter 5

Table J.1: **Weighted percent intra-country rescore agreement, by item and by country**

Item ID	Scale	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
B1Q1S1	D	96	95	96	92	92	100	97	96	96	0
B1Q2S1	D	96	94	99	99	96	100	97	98	98	0
B1Q3S1	D	99	95	100	96	98	100	99	99	99	0
B1Q4S1	Q	99	99	99	99	98	100	97	98	99	0
B1Q5S1	P	97	93	93	85	84	100	97	87	90	0
B1Q6S1	P	94	96	99	90	97	100	98	96	97	0
B1Q7S1	Q	96	99	99	95	100	100	98	99	99	0
B1Q8S1	D	97	98	100	93	98	99	98	99	99	0
B1Q9S1	Q	98	99	99	96	96	99	98	99	98	0
B1Q10S1	P	97	98	99	95	99	99	100	99	99	0
B1Q11S1	P	96	96	98	89	94	100	96	93	97	0
B1Q12S1	P	92	90	97	86	98	100	95	95	96	0
B1Q13S1	D	96	97	100	94	98	100	99	98	99	0
B1Q14S1	Q	98	99	100	97	100	100	99	100	100	0
B1Q15S1	Q	99	99	100	97	99	100	99	100	99	0
B2Q1S1	P	97	97	99	97	94	100	99	99	100	0
B2Q2S1	P	95	94	97	87	90	100	99	91	92	0
B2Q3S1	P	96	97	100	99	98	100	99	97	99	0
B2Q4S1	Q	96	98	99	99	97	100	98	100	99	0
B2Q5S1	Q	98	98	100	95	97	100	98	98	100	0
B2Q6S1	P	92	94	98	91	92	100	96	94	94	0
B2Q7S1	P	91	93	96	92	92	99	93	92	96	0
B2Q8S1	D	96	95	97	91	94	98	98	93	92	0
B2Q9S1	Q	97	98	100	99	97	100	98	100	99	0
B2Q10S1	D	99	98	100	97	97	99	100	99	99	0
B2Q111S1	D	98	99	99	100	100	100	98	99	100	0
B2Q112S1	D	92	94	98	94	96	98	97	95	98	0
B2Q113S1	D	99	99	99	95	98	99	99	99	99	0
B2Q114S1	D	98	100	99	99	99	99	99	99	99	0
B2Q115S1	D	98	98	98	95	98	99	98	98	99	0
B3Q1S1	Q	99	98	100	98	98	100	99	99	99	0
B3Q2S1	D	94	95	98	80	91	100	95	93	96	0
B3Q3S1	Q	99	100	100	99	100	100	99	99	100	0
B3Q4S1	D	98	100	100	98	99	100	100	100	99	0
B3Q5S1	D	97	98	100	96	97	99	99	100	99	0
B3Q6S1	Q	100	100	100	99	99	99	99	99	99	0
B3Q7S1	P	92	98	100	82	95	100	95	96	98	0
B3Q8S1	P	97	98	100	82	84	100	87	93	90	0
B3Q9S1	P	91	95	97	91	94	100	95	95	95	0
B3Q10S1	D	92	96	98	91	100	100	95	86	99	0
B3Q11S1	P	97	95	99	90	97	100	98	98	98	0
B3Q12S1	P	99	96	98	96	97	100	99	99	99	0
B3Q13S1	P	95	95	100	93	99	100	98	98	98	0
B3Q14S1	Q	97	98	100	96	99	100	99	98	98	0
B3Q15S1	P	93	92	100	90	94	99	96	95	98	0
B4Q1S1	P	99	99	100	98	98	100	100	99	99	0
B4Q2S1	P	94	96	98	95	96	99	99	93	97	0
B4Q3S1	Q	98	99	100	98	99	97	99	99	99	0
B4Q4S1	D	96	95	95	88	95	99	96	94	97	0
B4Q5_1S1	D	95	93	98	89	98	100	96	97	98	0
B4Q5_2S1	Q	96	92	100	90	97	100	96	97	97	0
B4Q6S1	P	95	92	98	95	89	100	98	88	94	0
B4Q7S1	P	94	95	98	90	91	100	86	94	94	0
B4Q8S1	P	95	90	98	90	98	98	99	88	86	0
B4Q9S1	Q	98	98	98	96	97	99	95	99	97	0
B4Q10S1	Q	97	97	98	94	98	99	95	99	98	0
B4Q11S1	Q	98	98	100	95	98	99	98	99	98	0

BEST COPY AVAILABLE

Table J.1: **Weighted percent intra-country rescore agreement, by item and by country** (Concluded)

Item ID	Scale	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
B4Q121S1	D	98	98	98	98	99	100	98	98	98	0
B4Q122S1	D	95	97	99	94	98	100	99	98	98	0
B4Q123S1	D	95	97	99	96	97	100	100	98	98	0
B4Q124S1	D	98	98	99	98	100	100	100	98	98	0
B4Q125S1	Q	96	96	98	97	96	100	99	98	98	0
B4Q126S1	Q	97	96	99	94	98	99	100	98	99	0
B5Q1S1	P	96	98	100	93	98	100	100	100	99	0
B5Q2S1	P	98	98	100	88	95	100	100	99	99	0
B5Q3S1	P	96	99	100	95	96	100	98	98	99	0
B5Q4S1	P	98	98	98	83	94	98	93	86	93	0
B5Q5S1	P	94	93	100	91	94	99	98	97	98	0
B5Q6S1	P	91	94	98	79	92	99	94	95	90	0
B5Q7S1	D	99	98	100	98	99	100	99	99	100	0
B5Q8S1	D	95	98	98	88	94	98	95	95	97	0
B5Q9S1	Q	98	98	98	96	98	99	98	99	99	0
B5Q10S1	D	94	92	99	85	90	100	95	88	95	0
B5Q111S1	Q	97	97	100	68	97	100	95	94	98	0
B5Q112S1	Q	97	96	100	93	98	100	97	99	99	0
B5Q113S1	Q	96	95	100	88	98	100	97	98	98	0
B5Q114S1	D	95	96	98	91	98	100	97	98	98	0
B5Q12S1	D	98	98	98	92	99	100	99	98	99	0
B5Q13S1	Q	98	99	100	90	98	100	97	98	99	0
B5Q14S1	Q	97	100	100	95	98	99	99	97	98	0
B6Q1S1	P	99	98	98	97	99	100	99	99	98	0
B6Q2S1	Q	99	100	100	98	99	100	100	99	100	0
B6Q3S1	Q	99	98	100	99	99	100	100	98	99	0
B6Q4S1	D	99	100	99	98	100	100	100	100	99	0
B6Q5S1	Q	99	98	100	97	96	99	97	99	98	0
B6Q6S1	D	93	94	99	90	91	99	96	95	95	0
B6Q7S1	P	99	96	100	94	99	99	100	100	99	0
B6Q8S1	P	96	97	100	93	100	100	99	97	99	0
B6Q9S1	D	97	97	96	98	99	98	99	99	96	0
B6Q10S1	Q	94	94	100	97	98	100	99	97	98	0
B6Q11S1	D	99	98	99	98	98	99	99	100	99	0
B6Q12S1	D	99	98	100	98	100	92	99	100	98	0
B6Q13S1	D	98	97	100	98	99	100	99	99	99	0
B7Q1S1	D	99	99	99	98	99	100	99	100	99	0
B7Q2S1	Q	98	99	100	98	99	100	100	99	100	0
B7Q3S1	D	98	99	100	98	98	99	100	100	99	0
B7Q4S1	D	98	97	99	94	99	100	100	99	99	0
B7Q5S1	Q	97	98	99	94	99	100	97	99	99	0
B7Q6S1	Q	98	99	100	92	97	99	97	99	94	0
B7Q7S1	D	98	98	99	94	98	100	98	99	99	0
B7Q8S1	D	97	97	99	94	99	99	98	99	99	0
B7Q9S1	D	98	97	99	97	99	99	97	100	99	0
B7Q10S1	P	97	94	97	95	94	100	97	96	98	0
B7Q11S1	P	97	96	99	93	92	99	97	96	96	0
B7Q12S1	P	92	96	98	94	94	100	95	93	96	0
B7Q13S1	P	98	99	100	96	99	100	98	100	98	0
B7Q14S1	P	98	97	98	88	93	100	96	96	96	0
B7Q15S1	P	97	93	100	89	96	100	98	98	98	0
Average		**97**	**97**	**99**	**94**	**97**	**99**	**98**	**97**	**98**	**99**
Number of booklets		**688**	**313**	**504**	**268**	**599**	**589**	**301**	**1,187**	**1,143**	**1,986**

Table J.2: Weighted percent inter-country rescore agreement, by item and by country

Item ID	Scale	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
B1Q1S1	D	97	98	95	95	94	96	96	90	98	98
B1Q2S1	D	98	100	93	97	97	96	94	86	98	100
B1Q3S1	D	100	95	95	98	94	100	99	93	100	100
B1Q4S1	Q	98	98	97	99	97	98	98	100	98	100
B1Q5S1	P	83	87	94	85	79	90	96	83	84	91
B1Q6S1	P	97	95	95	92	99	96	97	93	100	97
B1Q7S1	Q	100	100	96	98	97	98	97	98	98	97
B1Q8S1	D	100	100	99	96	94	97	99	100	100	100
B1Q9S1	Q	100	98	98	99	100	97	97	98	95	97
B1Q10S1	P	98	100	97	97	99	98	100	98	98	95
B1Q11S1	P	100	95	97	92	91	92	96	93	90	98
B1Q12S1	P	97	91	98	93	93	92	94	95	96	96
B1Q13S1	D	98	100	98	92	100	100	99	93	100	98
B1Q14S1	Q	100	100	99	100	98	99	99	98	100	98
B1Q15S1	Q	100	100	98	96	99	98	99	98	100	100
B2Q1S1	P	95	98	96	98	94	99	99	96	100	98
B2Q2S1	P	91	84	96	81	87	86	94	90	92	95
B2Q3S1	P	94	100	100	97	97	97	99	96	98	97
B2Q4S1	Q	97	98	98	99	96	95	97	96	98	97
B2Q5S1	Q	98	96	98	97	97	94	98	98	100	100
B2Q6S1	P	88	94	91	85	94	92	96	87	93	95
B2Q7S1	P	92	100	93	85	89	95	90	81	87	92
B2Q8S1	D	94	86	94	75	91	93	97	93	86	91
B2Q9S1	Q	98	100	99	98	99	98	98	100	98	97
B2Q10S1	D	98	98	98	97	96	98	100	100	98	96
B2Q111S1	D	100	98	97	97	99	97	98	100	100	98
B2Q112S1	D	94	93	93	93	96	98	96	95	98	97
B2Q113S1	D	100	100	97	94	100	98	100	98	98	98
B2Q114S1	D	100	100	98	98	100	98	99	100	100	98
B2Q115S1	D	98	96	99	96	98	96	98	100	100	97
B3Q1S1	Q	100	100	96	99	98	100	99	98	100	98
B3Q2S1	D	86	93	89	83	94	92	94	93	95	85
B3Q3S1	Q	98	96	98	96	100	99	100	100	100	100
B3Q4S1	D	100	100	95	99	99	98	100	100	100	100
B3Q5S1	D	98	98	98	96	97	100	99	100	98	96
B3Q6S1	Q	100	98	98	99	97	97	99	100	100	97
B3Q7S1	P	95	96	94	84	96	95	94	98	96	95
B3Q8S1	P	82	80	91	74	77	82	87	86	86	87
B3Q9S1	P	95	95	96	88	91	94	94	92	95	96
B3Q10S1	D	98	100	97	93	92	98	94	100	100	100
B3Q11S1	P	97	95	96	90	94	97	98	95	92	95
B3Q12S1	P	100	100	96	90	97	100	98	98	100	100
B3Q13S1	P	100	98	96	94	99	98	97	95	98	100
B3Q14S1	Q	98	100	97	96	99	100	99	98	98	96
B3Q15S1	P	98	91	83	90	94	96	95	90	96	93
B4Q1S1	P	100	100	98	99	100	99	100	100	100	100
B4Q2S1	P	94	91	96	91	97	96	98	96	95	95
B4Q3S1	Q	98	100	97	93	100	93	99	100	98	100
B4Q4S1	D	88	93	92	93	90	94	96	92	92	90
B4Q5_1S1	D	98	98	97	94	97	97	96	96	93	96
B4Q5_2S1	Q	97	96	99	93	94	98	97	96	96	95
B4Q6S1	P	100	88	97	94	94	92	97	86	93	98
B4Q7S1	P	98	96	98	83	93	93	88	89	95	95
B4Q8S1	P	98	86	96	91	97	90	88	66	63	94
B4Q9S1	Q	98	98	98	99	99	99	95	100	95	96
B4Q10S1	Q	100	100	96	97	96	99	95	98	93	94
B4Q11S1	Q	98	100	99	97	98	100	97	100	98	97
B4Q121S1	D	98	98	97	97	99	98	97	100	96	96
B4Q122S1	D	98	100	100	93	98	95	99	100	96	94
B4Q123S1	D	100	96	98	94	99	98	99	95	92	96
B4Q124S1	D	100	100	98	98	100	100	100	96	98	100
B4Q125S1	Q	100	100	98	95	98	96	98	98	96	98
B4Q126S1	Q	95	100	98	97	100	96	100	98	98	98

BEST COPY AVAILABLE

Table J.2: **Weighted percent inter-country rescore agreement, by item and by country** (Concluded)

Item ID	Scale	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ French	Switzerland/ German	United States
B5Q1S1	P	100	80	97	97	97	100	100	100	100	100
B5Q2S1	P	100	100	99	91	95	99	99	100	100	100
B5Q3S1	P	95	100	98	95	95	95	97	100	100	100
B5Q4S1	P	92	79	95	82	88	88	92	93	83	94
B5Q5S1	P	95	93	97	87	96	93	98	90	95	96
B5Q6S1	P	94	77	95	83	90	91	93	66	68	98
B5Q7S1	D	100	100	96	95	100	98	99	98	100	98
B5Q8S1	D	94	91	100	88	97	93	93	98	84	97
B5Q9S1	Q	100	100	96	96	99	96	96	100	98	100
B5Q10S1	D	86	83	93	93	90	94	94	83	89	91
B5Q111S1	Q	97	100	97	90	97	100	95	96	80	98
B5Q112S1	Q	98	100	97	97	98	99	96	98	98	98
B5Q113S1	Q	98	100	97	96	99	98	96	98	95	98
B5Q114S1	D	95	100	96	95	98	97	96	96	90	100
B5Q12S1	D	100	100	98	93	97	100	99	96	100	100
B5Q13S1	Q	97	100	99	92	99	100	96	98	100	98
B5Q14S1	Q	98	95	98	95	99	97	99	96	96	97
B6Q1S1	P	100	100	98	97	97	97	99	100	90	100
B6Q2S1	Q	100	100	97	99	98	100	100	96	100	98
B6Q3S1	Q	100	100	96	99	99	100	100	100	100	98
B6Q4S1	D	97	100	98	99	100	99	100	100	100	100
B6Q5S1	Q	100	98	97	95	98	97	97	98	96	100
B6Q6S1	D	95	89	92	92	86	96	94	74	93	94
B6Q7S1	P	98	98	97	99	99	96	100	98	100	98
B6Q8S1	P	95	96	97	96	75	93	98	100	100	94
B6Q9S1	D	95	96	97	98	100	98	98	95	100	100
B6Q10S1	Q	97	100	95	96	100	100	99	100	98	98
B6Q11S1	D	95	100	95	98	100	100	98	100	100	100
B6Q12S1	D	98	98	96	99	97	93	99	100	96	98
B6Q13S1	D	97	100	96	98	97	97	99	100	98	100
B7Q1S1	D	100	100	98	99	100	99	99	100	100	98
B7Q2S1	Q	100	100	97	99	98	98	100	100	100	98
B7Q3S1	D	98	100	98	97	97	99	100	100	100	97
B7Q4S1	D	96	100	98	97	98	99	100	98	100	98
B7Q5S1	Q	100	98	97	95	99	100	97	100	98	97
B7Q6S1	Q	96	96	99	96	97	97	96	100	95	98
B7Q7S1	D	98	100	98	94	99	96	98	100	100	98
B7Q8S1	D	96	100	96	98	100	99	98	89	98	97
B7Q9S1	D	96	100	95	98	99	100	96	98	98	100
B7Q10S1	P	95	100	97	90	95	95	96	95	98	97
B7Q11S1	P	93	92	97	95	93	97	93	100	61	90
B7Q12S1	P	92	96	96	89	93	96	91	92	98	98
B7Q13S1	P	100	100	98	99	100	98	97	98	98	100
B7Q14S1	P	92	98	95	87	93	98	92	95	93	87
B7Q15S1	P	100	96	94	88	97	92	97	100	98	95
Average		**97**	**97**	**97**	**94**	**96**	**97**	**97**	**96**	**96**	**97**
Number of Booklets		**158**	**142**	**337**	**270**	**300**	**300**	**300**	**154**	**153**	**315**

Appendix K:

Additional Documentation

Chapter 10

Table K.1: **Response probabilities (.80), by item and by country**

Item ID	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
Prose scale										
COREQ1S1	190	224	190	190	123	190	190	190		190
B1Q5S1	318	318	318	318	318	318	389	318		318
B1Q6S1	297	297	297	297	297	297	297	297		297
B1Q10S1	248	248	248	248	248	248	248	248		248
B1Q11S1	377	377	377	377	377	377	377	464		377
B2Q1S1	254	254	254	254	254	254	254	254		254
B2Q3S1	230	230	230	230	230	230	230	230		230
B2Q6S1	329	329	329	329	329	329	329	329		329
B2Q7S1	374	374	374	374	402	374	374	374		374
B3Q7S1	306	306	306	306	327	306	306	306		306
B3Q8S1	338	338	338	338	338	338	338	338		338
B3Q9S1	279	253	287	287	287	287	287	339		287
B3Q11S1	281	281	281	281	281	281	281	281		281
B3Q12S1	318	318	318	318	318	318	318	391		318
B3Q13S1	297	297	297	297	297	297	297	258		297
B3Q15S1	246	295	246	246	246	246	246	246		246
B4Q1S1	188	188	188	188	188	188	188	188		188
B4Q2S1	313	298	298	298	292	298	298	298		276
B4Q6S1	314	314	314	314	265	314	314	314		236
B4Q7S1	306	306	306	306	306	306	306	306		336
B5Q1S1	192	192	192	192	192	192	192	192		192
B5Q2S1	226	226	226	226	226	226	264	226		226
B5Q3S1	255	255	255	255	255	255	255	255		255
B5Q4S1	350	350	350	350	350	350	350	350		350
B5Q5S1	324	324	324	324	324	324	324	324		324
B5Q6S1	293	316	316	316	316	316	331	316		316
B6Q1S1	209	209	209	209	209	209	209	209		209
B6Q7S1	275	275	275	275	275	275	275	275		275
B6Q8S1	310	310	310	310	310	310	310	270		275
B7Q10S1	271	271	271	271	271	271	271	271		271
B7Q11S1	349	349	349	349	494	349	349	383		349
B7Q13S1	206	206	206	206	206	206	206	206		206
B7Q14S1	294	294	294	294	294	294	267	294		294
B7Q15S1	275	275	275	275	275	275	293	275		228
Document scale										
COREQ2S1	182	182	182	182	182	182	182	182		182
B1Q1S1	260	291	291	291	291	291	291	291		291
B1Q2S1	254	254	254	254	254	254	254	254		254
B1Q13S1	315	246	237	237	237	237	237	237		237
B2Q8S1	322	322	322	365	322	322	322	269		280
B2Q10S1	304	304	304	304	304	304	347	266		304
B2Q111S1	231	231	231	231	231	231	231	231		231
B2Q112S1	280	280	280	280	280	280	280	280		236
B2Q113S1	227	227	227	227	227	227	227	227		227
B2Q114S1	221	221	221	221	221	221	221	221		243
B2Q115S1	277	237	237	237	237	237	237	237		237
B3Q2S1	341	341	341	341	341	341	341	341		341
B3Q5S1	258	258	258	258	258	296	258	258		258
B4Q4S1	321	321	321	321	321	321	321	321		321
B4Q5_1S1	294	294	294	294	294	294	272	273		322
B4Q121S1	229	229	229	229	229	229	229	229		229
B4Q122S1	256	256	256	256	285	256	228	256		256
B4Q123S1	222	222	222	222	222	222	222	222		222
B4Q124S1	195	195	195	195	195	195	195	195		195
B5Q7S1	242	242	242	242	242	242	242	242		242
B5Q8S1	291	291	291	291	248	291	259	361		291
B5Q10S1	295	295	295	295	295	295	295	295		295
B5Q114S1	325	332	302	302	302	302	302	302		302
B5Q12S1	364	313	313	313	313	313	313	313		313

BEST COPY AVAILABLE

Table K.1: **Response probabilities (.80), by item and by country** (Concluded)

Item ID	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
B6Q4S1	218	218	218	218	218	218	218	218		218
B6Q6S1	261	286	286	286	286	250	286	286		286
B6Q9S1	270	270	270	270	270	270	270	270		270
B6Q11S1	297	297	297	297	282	297	297	297		297
B7Q1S1	188	188	188	188	188	188	188	188		188
B7Q3S1	234	234	234	234	234	234	234	234		234
B7Q4S1	270	270	270	270	270	270	270	270		270
B7Q7S1	327	327	327	327	279	327	327	327		327
B7Q8S1	287	287	287	287	287	287	287	287		287
B7Q9S1	408	408	334	408	408	408	408	408		355
Quantitative scale										
COREQ3S1	262	262	262	262	262	262	262	281		262
COREQ5S1	232	180	232	232	232	232	232	232		232
B1Q4S1	289	289	289	289	289	289	260	289		289
B1Q7S1	300	300	300	300	300	300	300	300		300
B1Q9S1	302	302	302	302	302	302	302	302		302
B1Q14S1	293	293	293	293	293	327	293	293		293
B1Q15S1	265	265	265	265	265	265	265	265		265
B2Q4S1	315	315	315	315	315	315	315	315		315
B2Q5S1	408	408	408	408	408	408	408	408		408
B2Q9S1	255	255	255	255	255	255	255	255		287
B3Q1S1	276	276	276	276	276	276	276	276		276
B3Q3S1	277	277	277	277	277	277	277	277		277
B3Q6S1	294	273	273	273	273	308	273	273		273
B3Q14S1	328	328	328	328	353	328	328	328		328
B4Q3S1	272	272	272	272	272	272	272	272		272
B4Q5_2S1	302	302	302	302	302	302	302	302		337
B4Q9S1	324	324	324	324	324	324	324	324		324
B4Q10S1	381	381	381	381	381	381	381	381		381
B4Q11S1	280	280	280	280	280	280	280	280		307
B4Q125S1	229	229	229	229	229	229	229	229		229
B4Q126S1	225	225	225	225	225	225	225	225		225
B5Q9S1	293	293	293	293	293	293	293	293		293
B5Q111S1	336	336	336	385	336	336	336	336		322
B5Q112S1	331	331	331	331	331	331	331	331		331
B5Q13S1	335	335	335	335	335	335	335	335		335
B5Q14S1	308	308	367	308	341	308	308	308		308
B6Q2S1	311	280	280	280	280	315	280	280		280
B6Q3S1	253	286	253	253	253	253	253	253		253
B6Q5S1	287	287	287	287	287	287	287	287		287
B6Q10S1	348	348	348	348	348	348	348	348		348
B7Q2S1	268	268	268	268	268	268	268	268		268
B7Q5S1	317	258	317	317	317	317	317	317		317
B7Q6S1	321	321	500*	321	321	321	321	321		321

* Theoretical RP80 is above 500

Table K.2: **Mantel-Haenszel summary, by item and by country**

Item ID	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ German	Switzerland/ French	United States
COREQ1S1	—	8.4	2.0	—	0.5	2.7	5.3	1.1	1.5	0.0
COREQ2S1	1.4	1.4	1.6	1.0	1.1	0.5	2.6	1.4	0.6	0.0
COREQ3S1	0.5	5.1	2.4	0.2	7.7	—	0.9	3.5	2.3	0.0
COREQ4S1	—	—	—	—	—	—	—	—	—	—
COREQ5S1	5.7	—	1.1	3.0	7.7	4.4	0.3	2.5	1.4	0.0
COREQ6S1	—	—	—	—	—	—	—	—	—	—
B1Q1S1	—	2.7	4.2	3.6	3.6	1.9	9.7	3.3	2.0	0.0
B1Q2S1	10.7	2.5	0.9	4.8	8.1	4.2	3.7	4.6	0.9	0.0
B1Q3S1	—	—	—	—	—	—	—	—	—	—
B1Q4S1	1.4	2.3	5.8	1.2	—	0.6	5.7	0.8	2.2	0.0
B1Q5S1	3.1	6.8	6.7	3.6	—	14.0	6.5	1.7	1.6	0.0
B1Q6S1	9.4	3.9	2.7	3.2	8.7	8.3	1.8	1.3	6.0	0.0
B1Q7S1	3.4	1.2	2.9	0.6	6.5	17.3	3.0	6.4	0.8	0.0
B1Q8S1	—	—	—	—	—	—	—	—	—	—
B1Q9S1	1.5	3.1	1.9	1.3	2.8	2.4	1.2	1.2	2.6	0.0
B1Q10S1	2.2	2.9	3.3	2.3	13.9	3.2	6.0	5.7	1.4	0.0
B1Q11S1	2.1	3.8	3.3	3.1	3.1	—	3.8	10.3	3.5	0.0
B1Q12S1	—	—	—	—	—	—	—	—	—	—
B1Q13S1	—	—	0.4	3.4	10.3	5.6	6.2	10.7	1.4	0.0
B1Q14S1	21.6	5.3	8.1	4.6	1.0	1.7	19.9	0.4	7.8	—
B1Q15S1	2.8	2.8	5.5	6.7	4.3	1.9	3.8	2.0	1.4	0.0
B2Q1S1	3.2	6.1	5.8	4.7	1.8	10.6	1.4	4.8	12.6	0.0
B2Q2S1	—	—	—	—	—	—	—	—	—	—
B2Q3S1	2.9	0.0	0.7	2.5	1.3	1.5	2.8	1.5	3.2	0.0
B2Q4S1	2.6	2.6	0.4	0.7	0.9	0.4	0.2	3.4	3.2	0.0
B2Q5S1	0.4	2.3	1.7	1.9	6.2	0.5	5.7	6.3	0.6	0.0
B2Q6S1	9.5	0.5	—	4.2	2.9	0.4	3.0	2.0	3.9	0.0
B2Q7S1	2.7	0.2	39.6	—	1.2	9.0	18.5	3.2	9.2	0.0
B2Q8S1	2.0	4.1	20.6	12.2	0.8	—	—	24.8	—	0.0
B2Q9S1	4.8	1.5	2.3	2.4	3.8	0.8	—	1.9	0.9	0.0
B2Q10S1	0.6	0.6	4.7	1.3	—	—	9.2	1.3	0.4	0.0
B2Q111S1	18.6	4.5	1.1	1.8	5.2	12.0	10.1	8.4	5.7	0.0
B2Q112S1	28.0	3.1	3.0	2.4	5.6	1.0	—	1.4	2.2	0.0
B2Q113S1	23.7	8.3	2.1	28.0	0.3	7.5	1.1	12.5	0.9	0.0
B2Q114S1	12.8	6.6	5.5	1.0	6.1	6.7	—	13.3	0.6	0.0
B2Q115S1	—	4.2	2.7	3.2	7.0	1.1	1.6	19.3	3.9	0.0
B3Q1S1	2.3	1.0	4.9	1.5	4.2	3.9	3.5	3.0	1.5	0.0
B3Q2S1	3.4	13.0	4.4	6.1	7.8	4.9	10.3	20.3	9.5	0.0
B3Q3S1	3.1	2.9	10.1	4.8	4.2	1.1	3.8	2.6	2.1	0.0
B3Q4S1	—	—	—	—	—	—	—	—	—	—
B3Q5S1	19.5	5.0	1.2	6.2	5.8	14.2	8.3	8.0	9.7	—
B3Q6S1	—	17.4	3.6	13.7	2.2	1.0	4.2	18.8	5.9	—
B3Q7S1	3.2	2.8	—	—	0.7	5.4	5.8	1.8	2.5	0.0
B3Q8S1	7.1	1.9	2.0	5.4	2.9	2.4	0.5	1.8	7.9	0.0
B3Q9S1	—	—	—	5.4	2.3	1.6	2.1	3.2	1.8	0.0
B3Q10S1	—	—	—	—	—	—	—	—	—	—
B3Q11S1	1.1	3.2	2.0	2.1	10.8	3.1	1.2	2.0	1.6	0.0
B3Q12S1	3.5	1.6	12.0	14.9	8.1	—	2.9	8.8	1.0	0.0
B3Q13S1	1.4	8.5	1.0	0.8	4.9	—	5.4	7.8	4.0	0.0
B3Q14S1	0.2	0.8	8.2	—	0.7	0.7	2.1	10.2	4.7	0.0
B3Q15S1	3.0	—	8.2	1.0	5.4	4.9	3.6	2.2	5.2	0.0
B4Q1S1	0.2	2.0	0.2	1.9	1.9	0.7	4.2	2.5	1.7	0.0
B4Q2S1	—	1.8	3.7	—	13.0	12.1	—	1.6	13.0	0.0
B4Q3S1	3.0	1.0	0.2	3.9	2.7	0.1	2.2	0.8	0.9	0.0
B4Q4S1	1.7	3.2	4.6	2.1	5.9	3.0	0.8	8.0	8.9	0.0
B4Q5_1S1	1.6	1.9	1.3	1.8	—	—	—	1.8	2.0	0.0
B4Q5_2S1	0.8	2.0	1.6	4.0	1.8	4.9	—	3.2	1.8	0.0
B4Q6S1	13.1	6.1	24.7	—	1.4	16.0	—	9.8	18.2	0.0
B4Q7S1	3.9	2.2	2.1	3.3	7.7	2.9	—	3.6	10.4	0.0
B4Q8S1	—	—	—	—	—	—	—	—	—	—
B4Q9S1	17.7	4.4	3.7	9.3	2.2	3.6	2.0	3.8	1.2	0.0
B4Q10S1	1.4	1.6	1.9	23.4	11.4	0.5	7.8	2.0	3.8	0.0
B4Q11S1	12.4	3.1	1.0	1.1	2.4	2.5	—	1.8	1.5	0.0
B4Q121S1	2.4	1.7	2.6	2.5	1.8	12.2	1.0	15.9	1.0	0.0
B4Q122S1	1.8	3.5	5.7	—	—	4.1	3.0	8.3	2.3	0.0
B4Q123S1	3.1	3.3	2.5	0.4	2.5	6.3	1.6	11.0	0.9	0.0

Table K.2: **Mantel-Haenszel summary, by item and by country** (Concluded)

Item ID	Canada/ English	Canada/ French	France	Germany	Netherlands	Poland	Sweden	Switzerland/ German	Switzerland/ French	United States
B4Q124S1	4.7	1.8	10.3	3.3	3.5	5.3	0.3	21.1	1.5	0.0
B4Q125S1	0.7	2.6	5.4	1.4	1.4	6.0	0.9	12.3	1.4	0.0
B4Q126S1	0.8	2.3	2.6	7.3	11.8	1.3	2.9	15.9	1.2	0.0
B5Q1S1	1.0	0.1	1.9	11.6	6.3	1.1	7.5	0.2	0.9	0.0
B5Q2S1	2.4	0.0	1.9	2.4	—	0.4	11.5	4.6	4.3	0.0
B5Q3S1	4.6	2.0	2.2	2.4	3.1	6.1	1.5	1.7	2.7	0.0
B5Q4S1	3.9	2.7	0.8	10.5	18.4	0.7	1.3	11.0	10.7	0.0
B5Q5S1	5.4	13.3	1.5	8.3	1.9	4.3	7.4	3.2	10.0	0.0
B5Q6S1	—	0.9	1.5	8.5	—	0.8	0.5	9.6	3.2	0.0
B5Q7S1	5.3	0.0	5.6	2.0	2.5	2.9	4.7	2.2	1.6	0.0
B5Q8S1	2.0	10.0	4.1	—	—	—	2.8	1.6	7.1	0.0
B5Q9S1	4.4	0.2	1.1	9.1	2.6	2.1	3.8	2.8	1.0	0.0
B5Q10S1	7.2	6.7	17.0	3.1	2.1	6.5	7.2	4.7	1.3	0.0
B5Q111S1	0.8	1.3	9.8	5.0	0.7	2.2	—	4.3	—	0.0
B5Q112S1	25.4	0.3	—	0.6	3.5	1.1	2.8	6.5	1.6	0.0
B5Q113S1	—	—	—	—	—	—	—	—	—	—
B5Q114S1	—	—	23.9	0.5	1.0	6.0	4.5	1.9	3.5	0.0
B5Q12S1	—	0.4	13.7	0.8	3.6	9.4	0.1	2.3	2.4	0.0
B5Q13S1	5.2	3.9	2.1	1.5	1.4	0.1	0.6	3.7	1.9	0.0
B5Q14S1	3.1	1.2	1.8	—	10.8	10.9	3.4	—	2.0	0.0
B6Q1S1	4.1	1.3	4.2	1.2	5.6	0.7	3.3	0.8	2.4	0.0
B6Q2S1	—	9.1	2.6	6.9	—	13.5	1.3	6.3	2.9	0.0
B6Q3S1	12.2	—	2.6	0.7	9.4	3.8	6.6	5.8	1.0	0.0
B6Q4S1	1.5	5.0	3.1	1.7	1.6	4.7	0.6	5.2	3.2	0.0
B6Q5S1	1.2	2.3	1.9	1.3	2.9	1.8	0.1	6.2	1.6	0.0
B6Q6S1	—	8.1	—	4.6	3.8	11.4	2.5	10.3	2.9	—
B6Q7S1	2.8	0.5	3.0	3.3	13.9	13.7	2.1	5.6	2.8	0.0
B6Q8S1	16.5	8.4	7.2	2.3	6.9	—	—	2.1	11.8	0.0
B6Q9S1	2.9	5.0	—	1.4	1.3	6.5	6.6	2.3	7.4	0.0
B6Q10S1	4.8	2.3	0.6	1.8	2.3	10.5	7.8	3.7	1.7	0.0
B6Q11S1	0.3	4.5	4.2	—	7.7	2.3	1.0	2.1	3.7	0.0
B6Q12S1	—	—	—	—	—	—	—	—	—	—
B6Q13S1	—	—	—	—	—	—	—	—	—	—
B7Q1S1	1.0	1.4	1.9	3.9	4.1	0.4	0.4	0.8	0.8	0.0
B7Q2S1	15.7	3.5	5.8	2.7	3.0	0.4	11.0	13.2	2.2	0.0
B7Q3S1	12.5	2.2	1.8	3.4	6.5	0.1	5.6	1.7	0.7	0.0
B7Q4S1	1.2	9.2	3.1	12.4	11.6	16.3	1.2	2.3	2.4	0.0
B7Q5S1	3.9	—	10.4	1.4	17.2	1.4	3.4	0.8	0.5	0.0
B7Q6S1	16.2	1.1	5.0	1.4	1.2	5.3	8.2	—	3.2	0.0
B7Q7S1	0.8	1.6	10.4	—	29.5	2.8	9.7	11.4	1.8	0.0
B7Q8S1	2.4	1.1	7.4	11.5	1.3	3.0	3.3	13.4	1.0	0.0
B7Q9S1	5.0	2.2	13.2	1.6	18.9	1.8	—	—	2.7	0.0
B7Q10S1	0.9	2.2	0.4	5.9	10.5	2.0	32.2	1.7	6.4	0.0
B7Q11S1	35.9	29.1	4.4	—	4.8	—	1.8	7.1	1.2	0.0
B7Q12S1	—	—	—	—	—	—	—	—	—	—
B7Q13S1	1.8	0.2	0.7	1.2	15.8	0.6	0.2	19.4	0.2	0.0
B7Q14S1	2.9	1.4	4.4	13.5	—	0.2	2.1	12.7	14.4	0.0
B7Q15S1	14.0	3.8	4.8	10.4	—	2.6	—	28.8	0.2	0.0
Average	**5.9**	**3.7**	**4.9**	**4.5**	**5.5**	**4.5**	**4.4**	**6.1**	**3.5**	**0.0**
Number of items	**89**	**94**	**95**	**89**	**90**	**91**	**87**	**98**	**99**	**97**

Table K.3: **Item parameters, by item and by country**

	Canada/English		Canada/French		Switzerland/German		Switzerland/French		German	
Item ID	A	B	A	B	A	B	A	B	A	B
Prose scale										
COREQ1S1	0.719	-2.667	0.648	-2.129	0.719	-2.667	0.719	-2.667	0.602	-4.182
B1Q5S1	0.779	-0.104	0.779	-0.104	0.779	-0.104	0.779	-0.104	0.779	-0.104
B1Q6S1	0.773	-0.515	0.773	-0.515	0.773	-0.515	0.773	-0.515	0.773	-0.515
B1Q10S1	1.166	-1.100	1.166	-1.100	1.166	-1.100	1.166	-1.100	1.166	-1.100
B1Q11S1	0.701	0.921	0.701	0.921	0.701	0.921	0.701	0.921	0.701	0.921
B2Q1S1	1.143	-1.011	1.143	-1.011	1.143	-1.011	1.143	-1.011	1.143	-1.011
B2Q3S1	1.314	-1.373	1.314	-1.373	1.314	-1.373	1.314	-1.373	1.314	-1.373
B2Q6S1	0.766	0.095	0.766	0.095	0.766	0.095	0.766	0.095	0.766	0.095
B2Q7S1	0.857	1.071	0.857	1.071	0.857	1.071	0.857	1.071	0.899	1.663
B3Q7S1	1.035	-0.073	1.035	-0.073	1.035	-0.073	1.035	-0.073	1.049	0.341
B3Q8S1	0.734	0.218	0.734	0.218	0.734	0.218	0.734	0.218	0.734	0.218
B3Q9S1	0.733	-0.919	1.291	-0.938	1.247	-0.316	1.247	-0.316	1.247	-0.316
B3Q11S1	1.003	-0.576	1.003	-0.576	1.003	-0.576	1.003	-0.576	1.003	-0.576
B3Q12S1	0.624	-0.354	0.624	-0.354	0.624	-0.354	0.624	-0.354	0.624	-0.354
B3Q13S1	0.732	-0.584	0.732	-0.584	0.732	-0.584	0.732	-0.584	0.732	-0.584
B3Q15S1	0.718	-1.591	0.638	-0.777	0.718	-1.591	0.718	-1.591	0.718	-1.591
B4Q1S1	1.037	-2.356	1.037	-2.356	1.037	-2.356	1.037	-2.356	1.037	-2.356
B4Q2S1	0.907	-0.048	0.845	-0.415	0.845	-0.415	0.845	-0.415	0.914	-0.452
B4Q6S1	0.713	-0.275	0.713	-0.275	0.713	-0.275	0.713	-0.275	0.893	-0.985
B4Q7S1	0.941	-0.157	0.941	-0.157	0.941	-0.157	0.941	-0.157	0.941	-0.157
B5Q1S1	0.750	-2.573	0.750	-2.573	0.750	-2.573	0.750	-2.573	0.750	-2.573
B5Q2S1	0.892	-1.750	0.892	-1.750	0.892	-1.750	0.892	-1.750	0.892	-1.750
B5Q3S1	0.728	-1.389	0.728	-1.389	0.728	-1.389	0.728	-1.389	0.728	-1.389
B5Q4S1	0.976	0.732	0.976	0.732	0.976	0.732	0.976	0.732	0.976	0.732
B5Q5S1	0.759	-0.022	0.759	-0.022	0.759	-0.022	0.759	-0.022	0.759	-0.022
B5Q6S1	1.313	-0.161	1.130	0.186	1.130	0.186	1.130	0.186	1.130	0.186
B6Q1S1	0.965	-2.010	0.965	-2.010	0.965	-2.010	0.965	-2.010	0.965	-2.010
B6Q7S1	1.131	-0.617	1.131	-0.617	1.131	-0.617	1.131	-0.617	1.131	-0.617
B6Q8S1	0.914	-0.106	0.914	-0.106	0.914	-0.106	0.914	-0.106	0.914	-0.106
B7Q10S1	1.416	-0.533	1.416	-0.533	1.416	-0.533	1.416	-0.533	1.416	-0.533
B7Q11S1	0.956	0.693	0.956	0.693	0.956	0.693	0.956	0.693	0.584	2.947
B7Q13S1	1.150	-1.941	1.150	-1.941	1.150	-1.941	1.150	-1.941	1.150	-1.941
B7Q14S1	0.793	-0.540	0.793	-0.540	0.793	-0.540	0.793	-0.540	0.793	-0.540
B7Q15S1	0.897	-0.792	0.897	-0.792	0.897	-0.792	0.897	-0.792	0.897	-0.792

	United States		Netherlands		Poland		France		Sweden	
Item ID	A	B	A	B	A	B	A	B	A	B
Prose scale										
COREQ1S1	0.719	-2.667	0.719	-2.667	0.719	-2.667	0.719	-2.667	0.719	-2.667
B1Q5S1	0.779	-0.104	0.598	0.961	0.779	-0.104	0.779	-0.104	0.779	-0.104
B1Q6S1	0.773	-0.515	0.773	-0.515	0.773	-0.515	0.773	-0.515	0.773	-0.515
B1Q10S1	1.166	-1.100	1.166	-1.100	1.166	-1.100	1.166	-1.100	1.166	-1.100
B1Q11S1	0.701	0.921	0.701	0.921	0.504	2.147	0.701	0.921	0.701	0.921
B2Q1S1	1.143	-1.011	1.143	-1.011	1.143	-1.011	1.143	-1.011	1.143	-1.011
B2Q3S1	1.314	-1.373	1.314	-1.373	1.314	-1.373	1.314	-1.373	1.314	-1.373
B2Q6S1	0.766	0.095	0.766	0.095	0.766	0.095	0.654	1.138	0.766	0.095
B2Q7S1	0.857	1.071	0.857	1.071	0.857	1.071	0.857	1.071	0.857	1.071
B3Q7S1	1.035	-0.073	1.035	-0.073	1.035	-0.073	1.252	-0.495	1.035	-0.073
B3Q8S1	0.734	0.218	0.734	0.218	0.734	0.218	0.734	0.218	0.734	0.218
B3Q9S1	1.247	-0.316	1.247	-0.316	0.635	0.061	0.635	0.060	1.247	-0.316
B3Q11S1	1.003	-0.576	1.003	-0.576	1.003	-0.576	1.003	-0.576	1.003	-0.576
B3Q12S1	0.624	-0.354	0.624	-0.354	0.494	0.712	0.624	-0.354	0.624	-0.354
B3Q13S1	0.732	-0.584	0.732	-0.584	0.964	-1.058	0.732	-0.584	0.732	-0.584
B3Q15S1	0.718	-1.591	0.718	-1.591	0.718	-1.591	0.718	-1.591	0.718	-1.591
B4Q1S1	1.037	-2.356	1.037	-2.356	1.037	-2.356	1.037	-2.356	1.037	-2.356
B4Q2S1	0.845	-0.415	0.845	-0.415	0.845	-0.415	0.845	-0.415	1.024	-0.658
B4Q6S1	0.713	-0.275	0.713	-0.275	0.713	-0.275	0.713	-0.275	0.991	-1.472
B4Q7S1	0.941	-0.157	0.941	-0.157	0.941	-0.157	0.941	-0.157	0.802	0.277

Table K.3: **Item parameters, by item and by country** (Continued)

	United States		Netherlands		Poland		France		Sweden	
Item ID	A	B	A	B	A	B	A	B	A	B
Prose scale										
B5Q1S1	0.750	-2.573	0.750	-2.573	0.750	-2.573	0.750	-2.573	0.750	-2.573
B5Q2S1	0.892	-1.750	0.871	-1.035	0.892	-1.750	0.892	-1.750	0.892	-1.750
B5Q3S1	0.728	-1.389	0.728	-1.389	0.728	-1.389	0.728	-1.389	0.728	-1.389
B5Q4S1	0.976	0.732	0.976	0.732	0.976	0.732	0.976	0.732	0.976	0.732
B5Q5S1	0.759	-0.022	0.759	-0.022	0.759	-0.022	0.759	-0.022	0.759	-0.022
B5Q6S1	1.130	0.186	1.128	0.472	1.130	0.186	1.130	0.186	1.130	0.186
B6Q1S1	0.965	-2.010	0.965	-2.010	0.965	-2.010	0.965	-2.010	0.965	-2.010
B6Q7S1	1.131	-0.617	1.131	-0.617	1.131	-0.617	1.131	-0.617	1.131	-0.617
B6Q8S1	0.914	-0.106	0.914	-0.106	1.236	-0.635	0.914	-0.106	1.223	-0.550
B7Q10S1	1.416	-0.533	1.416	-0.533	1.416	-0.533	1.416	-0.533	1.416	-0.533
B7Q11S1	0.956	0.693	0.956	0.693	1.002	1.398	0.956	0.693	0.956	0.693
B7Q13S1	1.150	-1.941	1.150	-1.941	1.150	-1.941	1.150	-1.941	1.150	-1.941
B7Q14S1	0.793	-0.540	1.374	-0.638	0.793	-0.540	0.793	-0.540	0.793	-0.540
B7Q15S1	0.897	-0.792	1.417	-0.112	0.897	-0.792	0.897	-0.792	0.850	-1.760

	Canada/English		Canada/French		Switzerland/German		Switzerland/French		German	
Item ID	A	B	A	B	A	B	A	B	A	B
Document scale										
COREQ2S1	0.584	-2.451	0.584	-2.451	0.584	-2.451	0.584	-2.451	0.584	-2.451
B1Q1S1	0.495	-1.217	0.585	-0.374	0.585	-0.374	0.585	-0.374	0.585	-0.374
B1Q2S1	0.771	-0.740	0.771	-0.740	0.771	-0.740	0.771	-0.740	0.771	-0.740
B1Q13S1	0.336	-0.955	0.834	-0.814	0.593	-1.389	0.593	-1.389	0.593	-1.389
B2Q8S1	0.679	0.418	0.679	0.418	0.679	0.418	0.575	1.018	0.679	0.418
B2Q10S1	0.730	0.145	0.730	0.145	0.730	0.145	0.730	0.145	0.730	0.145
B2Q111S1	1.436	-0.693	1.436	-0.693	1.436	-0.693	1.436	-0.693	1.436	-0.693
B2Q112S1	1.020	0.015	1.020	0.015	1.020	0.015	1.020	0.015	1.020	0.015
B2Q113S1	1.403	-0.776	1.403	-0.776	1.403	-0.776	1.403	-0.776	1.403	-0.776
B2Q114S1	2.409	-0.646	2.409	-0.646	2.409	-0.646	2.409	-0.646	2.409	-0.646
B2Q115S1	0.472	-0.976	1.025	-0.814	1.025	-0.814	1.025	-0.814	1.025	-0.814
B3Q2S1	0.732	0.861	0.732	0.861	0.732	0.861	0.732	0.861	0.732	0.861
B3Q5S1	0.787	-0.639	0.787	-0.639	0.787	-0.639	0.787	-0.639	0.787	-0.639
B4Q4S1	1.088	0.849	1.088	0.849	1.088	0.849	1.088	0.849	1.088	0.849
B4Q5_1S1	1.316	0.460	1.316	0.460	1.316	0.460	1.316	0.460	1.316	0.460
B4Q121S1	0.746	-1.260	0.746	-1.260	0.746	-1.260	0.746	-1.260	0.746	-1.260
B4Q122S1	0.790	-0.681	0.790	-0.681	0.790	-0.681	0.790	-0.681	0.639	-0.369
B4Q123S1	0.851	-1.245	0.851	-1.245	0.851	-1.245	0.851	-1.245	0.851	-1.245
B4Q124S1	0.946	-1.675	0.946	-1.675	0.946	-1.675	0.946	-1.675	0.946	-1.675
B5Q7S1	1.034	-0.699	1.034	-0.699	1.034	-0.699	1.034	-0.699	1.034	-0.699
B5Q8S1	0.841	0.051	0.841	0.051	0.841	0.051	0.841	0.051	0.941	-0.667
B5Q10S1	0.978	0.254	0.978	0.254	0.978	0.254	0.978	0.254	0.978	0.254
B5Q114S1	1.008	0.863	1.085	1.042	0.910	0.340	0.910	0.340	0.910	0.340
B5Q12S1	0.587	1.017	0.804	0.431	0.804	0.431	0.804	0.431	0.804	0.431
B6Q4S1	0.986	-1.191	0.986	-1.191	0.986	-1.191	0.986	-1.191	0.986	-1.191
B6Q6S1	0.817	-0.549	0.787	-0.116	0.787	-0.116	0.787	-0.116	0.787	-0.116
B6Q9S1	0.979	-0.205	0.979	-0.205	0.979	-0.205	0.979	-0.205	0.979	-0.205
B6Q11S1	0.746	0.035	0.746	0.035	0.746	0.035	0.746	0.035	0.806	-0.172
B7Q1S1	1.002	-1.766	1.002	-1.766	1.002	-1.766	1.002	-1.766	1.002	-1.766
B7Q3S1	0.946	-0.920	0.946	-0.920	0.946	-0.920	0.946	-0.920	0.946	-0.920
B7Q4S1	0.843	-0.348	0.843	-0.348	0.843	-0.348	0.843	-0.348	0.843	-0.348
B7Q7S1	0.776	0.650	0.776	0.650	0.776	0.650	0.776	0.650	1.009	-0.025
B7Q8S1	1.236	0.284	1.236	0.284	1.236	0.284	1.236	0.284	1.236	0.284
B7Q9S1	0.936	2.373	0.936	2.373	1.379	1.250	0.936	2.373	0.936	2.373

Table K.3: **Item parameters, by item and by country** (Continued)

	United States		Netherlands		Poland		France		Sweden	
Item ID	A	B	A	B	A	B	A	B	A	B
Document scale										
COREQ2S1	0.584	-2.451	0.584	-2.451	0.584	-2.451	0.584	-2.451	0.584	-2.451
B1Q1S1	0.585	-0.374	0.585	-0.374	0.585	-0.374	0.585	-0.374	0.585	-0.374
B1Q2S1	0.771	-0.740	0.771	-0.740	0.771	-0.740	0.771	-0.740	0.771	-0.740
B1Q13S1	0.593	-1.389	0.593	-1.389	0.593	-1.389	0.593	-1.389	0.593	-1.389
B2Q8S1	0.679	0.418	0.679	0.418	0.559	-0.849	0.679	0.418	0.752	-0.277
B2Q10S1	0.730	0.145	0.610	0.758	0.698	-0.615	0.730	0.145	0.730	0.145
B2Q111S1	1.436	-0.693	1.436	-0.693	1.436	-0.693	1.436	-0.693	1.436	-0.693
B2Q112S1	1.020	0.015	1.020	0.015	1.020	0.015	1.020	0.015	0.938	-0.897
B2Q113S1	1.403	-0.776	1.403	-0.776	1.403	-0.776	1.403	-0.776	1.403	-0.776
B2Q114S1	2.409	-0.646	2.409	-0.646	2.409	-0.646	2.409	-0.646	0.724	-1.023
B2Q115S1	1.025	-0.814	1.025	-0.814	1.025	-0.814	1.025	-0.814	1.025	-0.814
B3Q2S1	0.732	0.861	0.732	0.861	0.732	0.861	0.732	0.861	0.732	0.861
B3Q5S1	0.708	-0.030	0.787	-0.639	0.787	-0.639	0.787	-0.639	0.787	-0.639
B4Q4S1	1.088	0.849	1.088	0.849	1.088	0.849	1.088	0.849	1.088	0.849
B4Q5_1S1	1.316	0.460	1.802	0.197	1.275	0.036	1.316	0.460	1.501	1.064
B4Q121S1	0.746	-1.260	0.746	-1.260	0.746	-1.260	0.746	-1.260	0.746	-1.260
B4Q122S1	0.790	-0.681	1.091	-0.919	0.790	-0.681	0.790	-0.681	0.790	-0.681
B4Q123S1	0.851	-1.245	0.851	-1.245	0.851	-1.245	0.851	-1.245	0.851	-1.245
B4Q124S1	0.946	-1.675	0.946	-1.675	0.946	-1.675	0.946	-1.675	0.946	-1.675
B5Q7S1	1.034	-0.699	1.034	-0.699	1.034	-0.699	1.034	-0.699	1.034	-0.699
B5Q8S1	0.841	0.051	0.866	-0.527	0.606	1.006	0.841	0.051	0.841	0.051
B5Q10S1	0.978	0.254	0.978	0.254	0.978	0.254	0.978	0.254	0.978	0.254
B5Q114S1	0.910	0.340	0.910	0.340	0.910	0.340	0.910	0.340	0.910	0.340
B5Q12S1	0.804	0.431	0.804	0.431	0.804	0.431	0.804	0.431	0.804	0.431
B6Q4S1	0.986	-1.191	0.986	-1.191	0.986	-1.191	0.986	-1.191	0.986	-1.191
B6Q6S1	1.302	-0.397	0.787	-0.116	0.787	-0.116	0.841	0.403	0.787	-0.116
B6Q9S1	0.979	-0.205	0.979	-0.205	0.979	-0.205	0.878	0.924	0.979	-0.205
B6Q11S1	0.746	0.035	0.746	0.035	0.746	0.035	0.746	0.035	0.746	0.035
B7Q1S1	1.002	-1.766	1.002	-1.766	1.002	-1.766	1.002	-1.766	1.002	-1.766
B7Q3S1	0.946	-0.920	0.946	-0.920	0.946	-0.920	0.946	-0.920	0.946	-0.920
B7Q4S1	0.843	-0.348	0.843	-0.348	0.843	-0.348	0.843	-0.348	0.843	-0.348
B7Q7S1	0.776	0.650	0.776	0.650	0.776	0.650	0.776	0.650	0.776	0.650
B7Q8S1	1.236	0.284	1.236	0.284	1.236	0.284	1.236	0.284	1.236	0.284
B7Q9S1	0.936	2.373	0.936	2.373	0.936	2.373	0.936	2.373	1.009	1.430

	Canada/English		Canada/French		Switzerland/German		Switzerland/French		German	
Item ID	A	B	A	B	A	B	A	B	A	B
Quantitative scale										
COREQ3S1	0.718	-1.416	0.718	-1.416	0.718	-1.416	0.718	-1.416	0.718	-1.416
COREQ5S1	0.689	-2.011	0.506	-3.387	0.689	-2.011	0.689	-2.011	0.689	-2.011
B1Q4S1	0.770	-0.837	0.770	-0.837	0.770	-0.837	0.770	-0.837	0.770	-0.837
B1Q7S1	1.055	-0.349	1.055	-0.349	1.055	-0.349	1.055	-0.349	1.055	-0.349
B1Q9S1	0.922	-0.425	0.922	-0.425	0.922	-0.425	0.922	-0.425	0.922	-0.425
B1Q14S1	0.940	-0.576	0.940	-0.576	0.940	-0.576	0.940	-0.576	0.940	-0.576
B1Q15S1	0.825	-1.210	0.825	-1.210	0.825	-1.210	0.825	-1.210	0.825	-1.210
B2Q4S1	0.666	-0.515	0.666	-0.515	0.666	-0.515	0.666	-0.515	0.666	-0.515
B2Q5S1	0.678	1.201	0.678	1.201	0.678	1.201	0.678	1.201	0.678	1.201
B2Q9S1	0.999	-1.219	0.999	-1.219	0.999	-1.219	0.999	-1.219	0.999	-1.219
B3Q1S1	1.153	-0.718	1.153	-0.718	1.153	-0.718	1.153	-0.718	1.153	-0.718
B3Q3S1	1.256	-0.644	1.256	-0.644	1.256	-0.644	1.256	-0.644	1.256	-0.644
B3Q6S1	1.153	-0.395	0.931	-0.952	0.931	-0.952	0.931	-0.952	0.931	-0.952
B3Q14S1	1.019	0.133	1.019	0.133	1.019	0.133	1.019	0.133	0.994	0.573
B4Q3S1	0.762	-1.154	0.762	-1.154	0.762	-1.154	0.762	-1.154	0.762	-1.154
B4Q5_2S1	1.141	-0.258	1.141	-0.258	1.141	-0.258	1.141	-0.258	1.141	-0.258
B4Q9S1	0.730	-0.250	0.730	-0.250	0.730	-0.250	0.730	-0.250	0.730	-0.250
B4Q10S1	1.377	1.322	1.377	1.322	1.377	1.322	1.377	1.322	1.377	1.322
B4Q11S1	0.785	-0.982	0.785	-0.982	0.785	-0.982	0.785	-0.982	0.785	-0.982
B4Q125S1	0.779	-1.928	0.779	-1.928	0.779	-1.928	0.779	-1.928	0.779	-1.928
B4Q126S1	0.873	-1.892	0.873	-1.892	0.873	-1.892	0.873	-1.892	0.873	-1.892

BEST COPY AVAILABLE

Table K.3: **Item parameters, by item and by country** (Concluded)

	Canada/English		Canada/French		Switzerland/German		Switzerland/French		German	
Item ID	A	B	A	B	A	B	A	B	A	B
Quantitative scale										
B5Q9S1	0.748	-0.800	0.748	-0.800	0.748	-0.800	0.748	-0.800	0.748	-0.800
B5Q111S1	0.934	0.206	0.934	0.206	0.934	0.206	0.730	0.875	0.934	0.206
B5Q112S1	1.130	0.276	1.130	0.276	1.130	0.276	1.130	0.276	1.130	0.276
B5Q13S1	1.134	0.353	1.134	0.353	1.134	0.353	1.134	0.353	1.134	0.353
B5Q14S1	1.111	-0.164	1.111	-0.164	0.721	0.520	1.111	-0.164	0.944	0.310
B6Q2S1	0.812	-0.378	1.029	-0.727	1.029	-0.727	1.029	-0.727	1.029	-0.727
B6Q3S1	0.898	-1.350	0.853	-0.786	0.898	-1.350	0.898	-1.350	0.898	-1.350
B6Q5S1	0.718	-0.956	0.718	-0.956	0.718	-0.956	0.718	-0.956	0.718	-0.956
B6Q10S1	0.841	0.333	0.841	0.333	0.841	0.333	0.841	0.333	0.841	0.333
B7Q2S1	1.063	-0.927	1.063	-0.927	1.063	-0.927	1.063	-0.927	1.063	-0.927
B7Q5S1	0.926	-0.136	1.515	-0.879	0.926	-0.136	0.926	-0.136	0.926	-0.136
B7Q6S1	0.832	-0.162	0.832	-0.162	0.417	3.593	0.832	-0.162	0.832	-0.162

	United States		Netherlands		Poland		France		Sweden	
Item ID	A	B	A	B	A	B	A	B	A	B
Quantitative scale										
COREQ3S1	0.718	-1.416	0.718	-1.416	0.782	-0.963	0.718	-1.416	0.718	-1.416
COREQ5S1	0.689	-2.011	0.689	-2.011	0.689	-2.011	0.689	-2.011	0.689	-2.011
B1Q4S1	0.770	-0.837	0.942	-1.168	0.770	-0.837	0.770	-0.837	0.770	-0.837
B1Q7S1	1.055	-0.349	1.055	-0.349	1.055	-0.349	1.055	-0.349	1.055	-0.349
B1Q9S1	0.922	-0.425	0.922	-0.425	0.922	-0.425	0.922	-0.425	0.922	-0.425
B1Q14S1	1.009	0.109	0.940	-0.576	0.940	-0.576	0.940	-0.576	0.940	-0.576
B1Q15S1	0.825	-1.210	0.825	-1.210	0.825	-1.210	0.825	-1.210	0.825	-1.210
B2Q4S1	0.666	-0.515	0.666	-0.515	0.666	-0.515	0.666	-0.515	0.666	-0.515
B2Q5S1	0.678	1.201	0.678	1.201	0.678	1.201	0.678	1.201	0.678	1.201
B2Q9S1	0.999	-1.219	0.999	-1.219	0.999	-1.219	0.999	-1.219	0.747	-0.911
B3Q1S1	1.153	-0.718	1.153	-0.718	1.153	-0.718	1.153	-0.718	1.153	-0.718
B3Q3S1	1.256	-0.644	1.256	-0.644	1.256	-0.644	1.256	-0.644	1.256	-0.644
B3Q6S1	0.945	-0.290	0.931	-0.952	0.931	-0.952	0.931	-0.952	0.931	-0.952
B3Q14S1	1.019	0.133	1.019	0.133	1.019	0.133	1.019	0.133	1.019	0.133
B4Q3S1	0.762	-1.154	0.762	-1.154	0.762	-1.154	0.762	-1.154	0.762	-1.154
B4Q5_2S1	1.141	-0.258	1.141	-0.258	1.141	-0.258	1.141	-0.258	0.959	0.249
B4Q9S1	0.730	-0.250	0.730	-0.250	0.730	-0.250	0.730	-0.250	0.730	-0.250
B4Q10S1	1.377	1.322	1.377	1.322	1.377	1.322	1.377	1.322	1.377	1.322
B4Q11S1	0.785	-0.982	0.785	-0.982	0.785	-0.982	0.785	-0.982	0.914	-0.342
B4Q125S1	0.779	-1.928	0.779	-1.928	0.779	-1.928	0.779	-1.928	0.779	-1.928
B4Q126S1	0.873	-1.892	0.873	-1.892	0.873	-1.892	0.873	-1.892	0.873	-1.892
B5Q9S1	0.748	-0.800	0.748	-0.800	0.748	-0.800	0.748	-0.800	0.748	-0.800
B5Q111S1	0.934	0.206	0.934	0.206	0.934	0.206	0.934	0.206	0.875	-0.109
B5Q112S1	1.130	0.276	1.130	0.276	1.130	0.276	1.304	-0.197	1.130	0.276
B5Q13S1	1.134	0.353	1.134	0.353	1.134	0.353	1.134	0.353	1.134	0.353
B5Q14S1	1.111	-0.164	1.111	-0.164	1.111	-0.164	1.111	-0.164	1.111	-0.164
B6Q2S1	0.935	-0.166	1.029	-0.727	1.029	-0.727	0.935	-0.166	1.029	-0.727
B6Q3S1	0.898	-1.350	0.898	-1.350	0.898	-1.350	0.898	-1.350	0.898	-1.350
B6Q5S1	0.718	-0.956	0.718	-0.956	0.718	-0.956	0.718	-0.956	0.718	-0.956
B6Q10S1	0.841	0.333	0.841	0.333	0.841	0.333	0.841	0.333	0.841	0.333
B7Q2S1	1.063	-0.927	1.063	-0.927	1.063	-0.927	1.063	-0.927	1.063	-0.927
B7Q5S1	0.926	-0.136	0.926	-0.136	0.926	-0.136	0.926	-0.136	0.926	-0.136
B7Q6S1	0.832	-0.162	0.832	-0.162	0.832	-0.162	0.832	-0.162	0.832	-0.162

Table K.4: **Item parameter treatments, by item and by country***

Item ID	Scale	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
COREQ1S1	P	*U*	C	C	C	*U*	C	C	C	C	C
COREQ2S1	D	C	C	C	C	C	C	C	C	C	C
COREQ3S1	Q	C	C	C	C	C	C	C	*U*	C	C
COREQ4S1	D	Drop	-	-	-	-	-	-	-	-	-
COREQ5S1	Q	C	*U*	C	C	C	C	C	C	C	C
COREQ6S1	P	Drop	-	-	-	-	-	-	-	-	-
B1Q1S1	D	*U*	C	C	C	C	C	C	C	C	C
B1Q2S1	D	C	C	C	C	C	C	C	C	C	C
B1Q3S1	D	Drop	-	-	-	-	-	-	-	-	-
B1Q4S1	Q	C	C	C	C	C	C	*U*	C	C	C
B1Q5S1	P	C	C	C	C	C	C	*U*	C	C	C
B1Q6S1	P	C	C	C	C	C	C	C	C	C	C
B1Q7S1	Q	C	C	C	C	C	C	C	C	C	C
B1Q8S1	D	Drop	-	-	-	-	-	-	-	-	-
B1Q9S1	Q	C	C	C	C	C	C	C	C	C	C
B1Q10S1	P	C	C	C	C	C	C	C	C	C	C
B1Q11S1	P	C	C	C	C	C	C	C	*U*	C	C
B1Q12S1	P	Drop	-	-	-	-	-	-	-	-	-
B1Q13S1	D	*U*	*U*	C	C	C	C	C	C	C	C
B1Q14S1	Q	C	C	C	C	C	*U*	C	C	C	C
B1Q15S1	Q	C	C	C	C	C	C	C	C	C	C
B2Q1S1	P	C	C	C	C	C	C	C	C	C	C
B2Q2S1	P	Drop	-	-	-	-	-	-	-	-	-
B2Q3S1	P	C	C	C	C	C	C	C	C	C	C
B2Q4S1	Q	C	C	C	C	C	C	C	C	C	C
B2Q5S1	Q	C	C	C	C	C	C	C	C	C	C
B2Q6S1	P	C	C	C	C	C	C	C	C	C	C
B2Q7S1	P	C	C	C	C	*U*	C	C	C	*U*	C
B2Q8S1	D	C	C	C	*U*	C	C	C	*U*	C	*U*
B2Q9S1	Q	C	C	C	C	C	C	C	C	C	*U*
B2Q10S1	D	C	C	C	C	C	C	*U*	*U*	C	C
B2Q111S1	D	C	C	C	C	C	C	C	C	C	C
B2Q112S1	D	C	C	C	C	C	C	C	C	C	*U*
B2Q113S1	D	C	C	C	C	C	C	C	C	C	C
B2Q114S1	D	C	C	C	C	C	C	C	C	C	*U*
B2Q115S1	D	*U*	C	C	C	C	C	C	C	C	C
B3Q1S1	Q	C	C	C	C	C	C	C	C	C	C
B3Q2S1	D	C	C	C	C	C	C	C	C	C	C
B3Q3S1	Q	C	C	C	C	C	C	C	C	C	C
B3Q4S1	D	Drop	-	-	-	-	-	-	-	-	-
B3Q5S1	D	C	C	C	C	C	*U*	C	C	C	C
B3Q6S1	Q	*U*	C	C	C	C	*U*	C	C	C	C
B3Q7S1	P	C	C	C	C	*U*	C	C	C	*U*	C
B3Q8S1	P	C	C	C	C	C	C	C	C	C	C
B3Q9S1	P	*U*	*U*	C	C	C	C	C	C	*U*	C
B3Q10S1	D	Drop	-	-	-	-	-	-	-	-	-
B3Q11S1	P	C	C	C	C	C	C	C	C	C	C
B3Q12S1	P	C	C	C	C	C	C	C	*U*	C	C
B3Q13S1	P	C	C	C	C	C	C	C	*U*	C	C
B3Q14S1	Q	C	C	C	C	*U*	C	C	C	C	C
B3Q15S1	P	C	*U*	C	C	C	C	C	C	C	C
B4Q1S1	P	C	C	C	C	C	C	C	C	C	C
B4Q2S1	P	*U*	C	C	C	*U*	C	C	C	C	*U*
B4Q3S1	Q	C	C	C	C	C	C	C	C	C	C
B4Q4S1	D	C	C	C	C	C	C	C	C	C	C
B4Q5_1S1	D	C	C	C	C	C	C	*U*	*U*	C	*U*
B4Q5_2S1	Q	C	C	C	C	C	C	C	C	C	*U*
B4Q6S1	P	C	C	C	C	*U*	C	C	C	C	*U*
B4Q7S1	P	C	C	C	C	C	C	C	C	C	*U*
B4Q8S1	P	Drop	-	-	-	-	-	-	-	-	-
B4Q9S1	Q	C	C	C	C	C	C	C	C	C	C
B4Q10S1	Q	C	C	C	C	C	C	C	C	C	C
B4Q11S1	Q	C	C	C	C	C	C	C	C	C	*U*
B4Q121S1	D	C	C	C	C	C	C	C	C	C	C

---

* In this table, *U* indicates that an item parameter was unique for a given country; C indicates that the item parameter was common to other countries. The word "drop" in the Canada/English column indicates that the item was dropped from the IRT analysis *for all countries*.

Table K.4: Item parameter treatments, by item and by country* (Concluded)

Item ID	Scale	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
B4Q122S1	D	C	C	C	C	*U*	C	*U*	C	C	C
B4Q123S1	D	C	C	C	C	C	C	C	C	C	C
B4Q124S1	D	C	C	C	C	C	C	C	C	C	C
B4Q125S1	Q	C	C	C	C	C	C	C	C	C	C
B4Q126S1	Q	C	C	C	C	C	C	C	C	C	C
B5Q1S1	P	C	C	C	C	C	C	C	C	C	C
B5Q2S1	P	C	C	C	C	C	C	*U*	C	C	C
B5Q3S1	P	C	C	C	C	C	C	C	C	C	C
B5Q4S1	P	C	C	C	C	C	C	C	C	C	C
B5Q5S1	P	C	C	C	C	C	C	C	C	C	C
B5Q6S1	P	*U*	C	C	C	C	C	*U*	C	C	C
B5Q7S1	D	C	C	C	C	C	C	C	C	C	C
B5Q8S1	D	C	C	C	C	*U*	C	*U*	*U*	C	C
B5Q9S1	Q	C	C	C	C	C	C	C	C	C	C
B5Q10S1	D	C	C	C	C	C	C	C	C	C	C
B5Q111S1	Q	C	C	C	*U*	C	C	C	C	C	*U*
B5Q112S1	Q	C	C	C	C	C	C	C	C	*U*	C
B5Q113S1	Q	Drop	-	-	-	-	-	-	-	-	-
B5Q114S1	D	*U*	*U*	C	C	C	C	C	C	C	C
B5Q12S1	D	*U*	C	C	C	C	C	C	C	C	C
B5Q13S1	Q	C	C	C	C	C	C	C	C	C	C
B5Q14S1	Q	C	C	*U*	C	*U*	C	C	C	C	C
B6Q1S1	P	C	C	C	C	C	C	C	C	C	C
B6Q2S1	Q	*U*	C	C	C	C	C	*U*	C	C	C
B6Q3S1	Q	C	*U*	C	C	C	C	C	C	C	C
B6Q4S1	D	C	C	C	C	C	C	C	C	C	C
B6Q5S1	Q	C	C	C	C	C	C	C	C	C	C
B6Q6S1	D	*U*	C	C	C	C	*U*	C	C	*U*	C
B6Q7S1	P	C	C	C	C	C	C	C	C	C	C
B6Q8S1	P	C	C	C	C	C	C	C	*U*	C	*U*
B6Q9S1	D	C	C	C	C	C	C	C	C	*U*	C
B6Q10S1	Q	C	C	C	C	C	C	C	C	C	C
B6Q11S1	D	C	C	C	C	*U*	C	C	C	C	C
B6Q12S1	D	Drop	-	-	-	-	-	-	-	-	-
B6Q13S1	D	Drop	-	-	-	-	-	-	-	-	-
B7Q1S1	D	C	C	C	C	C	C	C	C	C	C
B7Q2S1	Q	C	C	C	C	C	C	C	C	C	C
B7Q3S1	D	C	C	C	C	C	C	C	C	C	C
B7Q4S1	D	C	C	C	C	C	C	C	C	C	C
B7Q5S1	Q	C	*U*	C	C	C	C	C	C	C	C
B7Q6S1	Q	C	C	*U*	C	C	C	C	C	C	C
B7Q7S1	D	C	C	C	C	*U*	C	C	C	C	C
B7Q8S1	D	C	C	C	C	C	C	C	C	C	C
B7Q9S1	D	C	C	*U*	C	C	C	C	C	C	*U*
B7Q10S1	P	C	C	C	C	C	C	C	C	C	C
B7Q11S1	P	C	C	C	C	*U*	C	C	*U*	C	C
B7Q12S1	P	Drop	-	-	-	-	-	-	-	-	-
B7Q13S1	P	C	C	C	C	C	C	C	C	C	C
B7Q14S1	P	C	C	C	C	C	C	*U*	C	C	C
B7Q15S1	P	C	C	C	C	C	C	*U*	C	C	*U*
Total no. of common items		**89**	**94**	**98**	**99**	**89**	**97**	**90**	**91**	**95**	**87**

[*] In this table, *U* indicates that an item parameter was unique for a given country; C indicates that the item parameter was common to other countries. The word "drop" in the Canada/English column indicates that the item was dropped from the IRT analysis *for all countries*.

An Overview of IRT Scale-Linking Procedures

When the IRT model is true, the scales from separate estimation runs are linearly related. Let k_1 and k_2 be two arbitrarily defined constants, with k_2 greater than 0. Further, define $a_i^*=a_i/k_2$, $b_i^*=k_1+k_2b_i$, and $\theta_j^*=k_1+k_2\theta$. Then, $P(\theta,a_i,b_i)=P(\theta^*,a_i^*,b_i^*)$. Additional constraints must be imposed during parameter estimation for the model to be identified. If item parameters from separate estimation runs are known for two or more items, k_1 and k_2 can be determined exactly. In similar fashion, if θ and θ^* are known for two or more individuals, the linear relationship between the scales can again be determined exactly. In practice, however, *estimates* of the IRT item and person parameters are all that are available. Even when the model is true, the linear relationship between item and person estimates from separate runs is not exact. As used here, scale-linking procedures refer to ways of obtaining item parameter estimates for sets of items to ensure that they are on a common scale, or to approaches for estimating k_1 and k_2 using multiple sets of item parameter estimates or multiple estimates of the θ-distribution.

Scale-linking procedures are usually based on one of two general approaches, the *common-item* approach or the *common-population* approach. A brief overview of these two approaches is provided below. For simplicity's sake, the discussion assumes that there are two scales to be linked—say, scale 1 and scale 2.

Common-item linking

Common-item linking can be performed when the scales to be linked are based on item pools that share sets of items. One group of examinees (Group 1) is administered item sets *X* and *Y* while the other group (Group 2) is administered sets *Y* and *Z*. Item set *Y* contains the so-called "common items."

There are at least three common-item scale-linking methods: the *concurrent-calibration method*, the *sequential-calibration* method, and the *parallel-calibration* method.

The concurrent-calibration method estimates item parameters for all three item sets (*X*, *Y*, and *Z*) in a single estimation run using the data from both groups of examinees. The linear scale indeterminacy for this combined estimation run can be resolved by standardizing the estimated θ-distribution or the distribution of $\hat{b}_i$'s. The item parameter estimates for all three sets of items are automatically on the same scale since they were jointly obtained subject to whatever linear constraints were imposed to set the unit and origin of the θ scale for that estimation run.

In the sequential method, parameter estimates are obtained for item sets *X* and *Y* using only the data from Group 1. The estimates for item set *Y* are then treated as known and used with the data from Group 2 to obtain parameter estimates for item set *Z* in an separate estimation run. Constraining the item parameter values for item set *Y* in Group 2 resolves the linear indeterminacy so that estimates for item set *Z* are on the same scale as those for item sets *X* and *Y*.

In the parallel method, parameters for item sets *X* and *Y* are estimated using data from Group 1 only. In a separate calibration run, parameters for items sets *Z* and *Y* are estimated using data from Group 2 only. Therefore, two sets of item parameter estimates are obtained for item set *Y*, one based on Group 1 data and the other based on Group 2 data. These item parameter estimates are used to estimate the linear scale transformation coefficients which transform the θ-scale from the Group 2 estimation run to the metric of the Group 1 θ-scale.

A number of procedures have been proposed for obtaining estimates of k_1 and k_2 using the parallel method (Marco 1977.) Variants of the mean-and-sigma approach which incorporate information about the precision of *b* estimates and attempt to make the procedure more resistant to outlier estimates have been proposed by Linn, Levine, Hastings, and Wardrop (1981) and Stocking and Lord (1983).

Least square methods by Stocking and Lord attempt to use information from the estimates of both the *a* and *b* parameters. Stocking and Lord estimate k_l and k_2 by minimizing the function

$$F = \frac{1}{T} \sum_{j=1}^{T} \sum_{i=1}^{l} [P(\theta^*, a_i^*, b_i^*) - P(\theta, a_i, b_i)]^2$$

where *T* is a spaced sample of θ values and *l* is the number of common items. Slightly different procedures in this same tradition have been proposed by Haebara (1980) and Divgi (1985). Details on these procedures are beyond the scope of this report. Research suggests that, in general, the latter set of approaches perform better than the simple mean-and-sigma method (Way and Tang 1991).

Common-population linking

Common-population linking is appropriate when random samples from a population have been administered the item sets whose scales are to be linked. In the case of common-population linking, random sample 1 is administered item set *X* but not item set *Y*, and random sample 2 is administered item set *Y* but not item set *X*.

In discussing common-population linking methods, it is useful to distinguish two separate situations: 1) parameter estimates are already available for both item sets, but the estimates are not on the same scale (i.e., they have been estimated in separate calibration runs), and 2) parameter estimates do not exist for either set of items.

If item parameter estimates are already available for both item sets, a procedure analogous to the parallel approach of common-item linking can be employed. Separate estimates of the first two moments of the distribution of q are obtained based on the two different item sets (say, $\hat{\mu}_x$, $\hat{\mu}_y$ and $\hat{\sigma}_x$) and $\hat{\sigma}_y$). An estimate of the linear transformation that links scale *X* to scale *Y* can be obtained by setting the estimates of these first two moments as equal and adjusting the item parameters accordingly. More specifically:

$$\hat{k}_2 = \frac{\hat{\sigma}_x}{\hat{\sigma}_y}; \quad \hat{k}_1 = \hat{\mu}_y - \hat{k}_2 \hat{\mu}_x$$

If neither set of items has yet been calibrated, a procedure analogous to the concurrent common-item method can be used. Item parameter estimates for both sets of items are obtained in a single calibration run. The scale of the solution is set by standardizing the θ distribution for the combined sample 1 and sample 2 distribution.

Appendix L:

Additional Documentation

Chapter 11

Table L.1: Weighted percent of correct responses, by item and by country

Item ID	Scale	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
COREQ1S1	P	93	88	98	98	98	95	97	87	91	97
COREQ2S1	D	95	93	96	96	96	95	96	89	89	95
COREQ3S1	Q	82	82	88	88	85	84	82	61	73	85
COREQ4S1	D	95	93	98	96	97	97	87	86	84	95
COREQ5S1	Q	92	92	93	92	93	91	88	76	82	91
COREQ6S1	P	87	82	95	96	94	94	97	78	77	98
B1Q1S1	D	84	69	79	79	82	82	74	59	60	75
B1Q2S1	D	90	83	79	85	86	91	81	68	74	89
B1Q3S1	D	93	87	74	79	73	89	74	54	68	81
B1Q4S1	Q	79	68	78	76	80	72	84	63	63	79
B1Q5S1	P	61	60	52	53	59	60	35	48	43	66
B1Q6S1	P	70	64	63	75	66	62	72	61	50	77
B1Q7S1	Q	66	57	75	69	74	71	64	60	53	72
B1Q8S1	D	68	48	62	50	56	65	44	40	23	36
B1Q9S1	Q	69	63	73	72	77	69	68	50	55	72
B1Q10S1	P	88	86	82	88	89	88	89	70	68	87
B1Q11S1	P	35	29	42	26	37	36	31	13	23	48
B1Q12S1	P	50	39	29	20	36	48	23	10	28	52
B1Q13S1	D	75	84	92	88	88	81	92	80	80	93
B1Q14S1	Q	67	53	82	84	72	59	74	60	67	84
B1Q15S1	Q	83	75	93	88	92	87	88	69	71	85
B2Q1S1	P	89	72	86	74	80	85	82	75	73	87
B2Q2S1	P	34	16	36	19	42	25	43	18	05	30
B2Q3S1	P	92	88	92	88	91	91	89	79	79	95
B2Q4S1	Q	63	59	70	76	69	65	66	53	58	74
B2Q5S1	Q	30	22	26	26	31	32	23	17	24	39
B2Q6S1	P	65	49	47	52	49	58	51	38	20	63
B2Q7S1	P	35	21	16	30	15	40	28	21	22	30
B2Q8S1	D	68	60	50	47	56	69	67	70	43	84
B2Q9S1	Q	86	83	87	85	89	84	87	76	73	81
B2Q10S1	D	70	55	67	67	68	70	55	69	59	81
B2Q111S1	D	94	86	95	89	95	93	96	75	83	93
B2Q112S1	D	72	64	85	82	81	70	77	64	64	94
B2Q113S1	D	95	89	93	96	88	95	93	77	84	98
B2Q114S1	D	95	88	95	94	96	94	97	77	86	92
B2Q115S1	D	82	84	84	88	92	88	94	80	81	98
B3Q1S1	Q	81	77	84	80	81	76	75	69	61	81
B3Q2S1	D	57	36	65	44	58	45	56	39	33	59
B3Q3S1	Q	82	62	82	81	83	78	79	64	57	80
B3Q4S1	D	76	76	94	95	94	71	88	75	66	94
B3Q5S1	D	81	75	90	92	90	72	86	62	73	93
B3Q6S1	Q	73	66	88	88	85	67	79	68	66	87
B3Q7S1	P	67	52	60	57	44	67	60	37	54	70
B3Q8S1	P	54	43	51	59	50	50	49	33	40	62
B3Q9S1	P	80	83	68	70	68	70	68	47	40	75
B3Q10S1	D	14	17	04	04	14	11	25	07	05	42
B3Q11S1	P	78	76	78	74	74	73	68	56	57	83
B3Q12S1	P	71	53	69	62	68	66	56	31	54	67
B3Q13S1	P	77	66	63	63	70	76	74	71	55	75
B3Q14S1	Q	57	45	46	63	44	56	54	41	46	63
B3Q15S1	P	85	73	93	89	91	89	90	87	71	89
B4Q1S1	P	97	93	99	96	97	97	97	92	94	97
B4Q2S1	P	64	56	59	55	67	63	76	60	49	78
B4Q3S1	Q	86	76	76	83	81	84	82	67	73	78
B4Q4S1	D	57	44	54	44	54	53	51	35	39	55
B4Q5_1S1	D	71	55	61	64	64	67	76	55	48	52
B4Q5_2S1	Q	69	51	63	68	70	71	66	47	50	54
B4Q6S1	P	70	50	73	55	79	63	69	44	43	91
B4Q7S1	P	63	56	59	66	64	59	65	46	43	55
B4Q8S1	P	74	59	34	57	67	69	69	58	72	71
B4Q9S1	Q	70	56	56	62	57	69	64	42	56	63

Table L.1: **Weighted percent of correct responses, by item and by country** (Concluded)

Item ID	Scale	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
B4Q10S1	Q	21	13	16	12	26	23	13	09	09	21
B4Q11S1	Q	82	78	81	78	79	81	78	60	70	70
B4Q121S1	D	89	81	87	93	91	90	92	89	85	93
B4Q122S1	D	89	80	84	89	78	84	92	68	78	84
B4Q123S1	D	92	95	91	93	95	89	91	88	88	92
B4Q124S1	D	93	95	95	98	97	93	97	93	95	97
B4Q125S1	Q	92	87	91	92	93	88	90	88	90	91
B4Q126S1	Q	93	89	91	94	91	95	95	83	88	94
B5Q1S1	P	95	97	96	97	91	96	97	89	90	98
B5Q2S1	P	91	90	95	95	91	90	82	79	81	95
B5Q3S1	P	83	89	84	89	85	86	86	66	73	85
B5Q4S1	P	43	29	34	42	39	36	27	22	20	47
B5Q5S1	P	60	67	56	66	46	68	55	35	37	63
B5Q6S1	P	67	55	36	53	42	66	44	33	32	62
B5Q7S1	D	85	85	90	92	91	86	88	80	81	90
B5Q8S1	D	78	58	73	76	89	70	84	34	55	78
B5Q9S1	Q	72	67	77	74	71	79	78	61	65	81
B5Q10S1	D	74	56	63	68	74	78	68	56	45	78
B5Q111S1	Q	50	34	56	35	59	55	53	39	37	67
B5Q112S1	Q	47	33	51	53	54	52	50	33	52	63
B5Q113S1	Q	49	40	53	55	57	54	54	34	55	69
B5Q114S1	D	56	40	70	66	70	59	68	42	57	76
B5Q12S1	D	52	49	65	63	64	55	67	54	53	72
B5Q13S1	Q	53	40	50	47	48	44	52	33	35	60
B5Q14S1	Q	62	59	46	57	53	60	73	58	52	75
B6Q1S1	P	93	97	94	96	94	93	96	86	84	97
B6Q2S1	Q	64	63	83	80	84	62	78	66	59	80
B6Q3S1	Q	81	70	93	88	89	78	90	78	76	93
B6Q4S1	D	92	89	96	93	97	91	94	82	88	96
B6Q5S1	Q	76	71	84	81	80	74	73	66	65	82
B6Q6S1	D	82	81	81	81	73	86	72	52	50	84
B6Q7S1	P	75	75	79	74	76	76	81	52	55	83
B6Q8S1	P	64	47	64	50	65	62	66	59	38	81
B6Q9S1	D	77	78	77	85	81	78	79	65	37	88
B6Q10S1	Q	48	47	49	52	48	42	45	25	40	63
B6Q11S1	D	74	72	75	79	79	68	69	53	54	81
B6Q12S1	D	62	63	70	65	74	60	40	31	58	53
B6Q13S1	D	84	74	73	83	69	89	83	56	74	93
B7Q1S1	D	96	96	98	99	96	95	98	91	93	97
B7Q2S1	Q	74	77	91	87	86	74	82	65	69	87
B7Q3S1	D	87	84	92	90	94	90	92	76	80	94
B7Q4S1	D	77	83	75	82	77	82	85	57	72	83
B7Q5S1	Q	63	77	56	69	61	61	56	41	57	63
B7Q6S1	Q	59	58	09	64	60	67	59	36	53	67
B7Q7S1	D	55	61	68	56	78	60	50	36	52	61
B7Q8S1	D	69	61	81	71	67	72	71	47	59	74
B7Q9S1	D	17	18	39	15	20	17	12	08	06	43
B7Q10S1	P	78	73	77	66	78	82	79	58	56	73
B7Q11S1	P	49	45	35	32	08	53	29	10	21	41
B7Q12S1	P	49	68	68	65	66	55	60	29	48	58
B7Q13S1	P	94	91	98	96	96	95	98	89	88	97
B7Q14S1	P	74	73	64	59	60	69	79	58	60	80
B7Q15S1	P	84	81	64	76	69	85	65	63	67	94

Table L.2: Weighted percent of omitted responses, by item and by country

Item ID	Scale	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
COREQ1S1	P	3	5	0	1	1	3	1	6	2	1
COREQ2S1	D	1	1	1	1	1	2	1	4	2	1
COREQ3S1	Q	2	2	1	2	1	3	1	8	3	2
COREQ4S1	D	1	1	0	1	0	1	1	6	2	1
COREQ5S1	Q	0	1	0	1	0	1	1	6	1	1
COREQ6S1	P	6	8	16	5	9	10	6	32	8	15
B1Q1S1	D	0	3	1	1	2	2	2	7	6	3
B1Q2S1	D	0	4	3	2	2	2	2	11	5	4
B1Q3S1	D	0	3	1	1	2	2	2	7	4	3
B1Q4S1	Q	2	6	4	3	3	7	2	12	6	9
B1Q5S1	P	4	5	5	6	3	7	3	15	6	11
B1Q6S1	P	6	11	11	7	7	14	6	20	14	15
B1Q7S1	Q	3	8	4	1	2	7	6	16	10	11
B1Q8S1	D	3	11	6	3	3	7	6	17	9	13
B1Q9S1	Q	3	10	6	4	3	8	7	21	13	13
B1Q10S1	P	1	5	7	4	4	4	3	14	9	10
B1Q11S1	P	3	7	8	7	5	8	5	20	16	16
B1Q12S1	P	5	12	12	11	7	11	9	37	20	20
B1Q13S1	D	0	0	0	0	0	0	0	1	1	0
B1Q14S1	Q	1	2	1	1	1	2	0	11	1	2
B1Q15S1	Q	1	1	1	1	1	0	0	3	1	0
B2Q1S1	P	2	9	5	10	4	6	4	12	9	5
B2Q2S1	P	4	12	11	11	6	15	10	25	17	8
B2Q3S1	P	1	5	3	6	4	6	3	11	9	4
B2Q4S1	Q	1	3	2	2	2	3	2	7	5	5
B2Q5S1	Q	2	7	4	5	4	6	4	15	11	14
B2Q6S1	P	3	9	9	6	13	10	7	17	9	16
B2Q7S1	P	8	16	11	13	11	18	13	35	17	34
B2Q8S1	D	2	13	6	4	8	6	2	13	7	2
B2Q9S1	Q	3	6	2	3	2	6	3	10	7	6
B2Q10S1	D	2	5	2	2	2	5	2	10	5	4
B2Q111S1	D	0	1	8	0	0	8	5	20	0	0
B2Q112S1	D	1	12	12	11	0	11	9	37	0	1
B2Q113S1	D	0	0	0	0	1	0	0	1	1	1
B2Q114S1	D	1	2	1	1	0	2	0	11	1	4
B2Q115S1	D	1	1	1	1	1	0	0	3	1	0
B3Q1S1	Q	2	8	4	5	5	7	7	14	12	8
B3Q2S1	D	6	20	6	13	6	16	11	23	22	18
B3Q3S1	Q	3	13	5	5	5	6	7	14	13	10
B3Q4S1	D	2	8	1	0	1	9	2	12	5	1
B3Q5S1	D	2	7	1	1	1	11	2	14	5	1
B3Q6S1	Q	3	8	1	0	1	11	2	11	5	1
B3Q7S1	P	1	4	5	5	6	5	3	15	5	8
B3Q8S1	P	4	17	8	8	8	11	7	26	15	15
B3Q9S1	P	3	5	7	7	7	9	7	24	10	14
B3Q10S1	D	1	2	2	1	1	3	1	2	3	2
B3Q11S1	P	2	2	4	2	4	7	3	12	6	5
B3Q12S1	P	2	3	4	2	6	7	3	12	6	4
B3Q13S1	P	2	4	4	3	5	2	3	9	3	5
B3Q14S1	Q	1	1	1	2	1	3	5	10	6	6
B3Q15S1	P	1	1	1	1	1	0	0	3	1	0
B4Q1S1	P	0	3	0	0	0	2	1	3	1	2
B4Q2S1	P	1	9	1	1	2	5	1	10	3	4
B4Q3S1	Q	1	3	1	1	1	3	1	6	2	3
B4Q4S1	D	5	19	11	10	7	9	9	30	18	27
B4Q5_1S1	D	8	28	16	12	13	12	13	29	21	31
B4Q5_2S1	Q	8	28	16	12	13	11	13	29	21	32
B4Q6S1	P	2	14	3	5	3	3	2	12	7	6
B4Q7S1	P	2	11	5	4	5	4	3	16	10	15
B4Q8S1	P	1	10	3	3	4	3	2	10	6	6
B4Q9S1	Q	0	1	3	3	3	3	2	8	5	5
B4Q10S1	Q	6	11	15	8	10	12	14	26	17	27
B4Q11S1	Q	1	2	4	4	3	2	3	12	6	7
B4Q121S1	D	2	0	4	3	0	2	0	9	3	1
B4Q122S1	D	0	0	0	2	1	3	5	10	6	8
B4Q123S1	D	1	1	1	1	1	0	0	3	1	1

Table L.2: **Weighted percent of omitted responses, by item and by country** (Concluded)

Item ID	Scale	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
B4Q124S1	D	2	1	3	2	1	3	1	7	2	0
B4Q125S1	Q	0	0	0	0	0	0	0	0	0	3
B4Q126S1	Q	0	0	0	0	0	0	0	0	0	0
B5Q1S1	P	3	2	2	3	6	3	1	9	4	2
B5Q2S1	P	2	1	1	2	2	3	1	6	3	1
B5Q3S1	P	1	1	4	2	3	3	3	10	3	7
B5Q4S1	P	5	7	16	8	11	14	10	25	17	17
B5Q5S1	P	4	3	8	5	10	9	6	23	12	11
B5Q6S1	P	4	5	10	7	10	11	6	24	10	11
B5Q7S1	D	1	0	1	1	1	1	1	3	2	2
B5Q8S1	D	1	3	1	1	1	1	1	5	3	2
B5Q9S1	Q	1	3	1	1	1	1	1	6	3	3
B5Q10S1	D	4	11	6	6	6	9	4	18	13	11
B5Q111S1	Q	4	12	4	4	13	9	4	16	7	22
B5Q112S1	Q	4	12	4	4	5	9	4	16	7	11
B5Q113S1	Q	4	12	4	4	6	9	4	16	7	12
B5Q114S1	D	4	12	4	4	5	9	4	16	7	12
B5Q12S1	D	0	1	1	0	1	0	0	3	1	0
B5Q13S1	Q	1	1	2	2	2	2	1	7	2	2
B5Q14S1	Q	0	0	0	0	0	0	0	0	0	3
B6Q1S1	P	1	1	1	0	2	2	0	4	3	2
B6Q2S1	Q	0	1	1	1	1	1	0	3	4	2
B6Q3S1	Q	0	3	1	0	1	1	1	3	2	1
B6Q4S1	D	2	1	2	1	1	4	1	7	5	3
B6Q5S1	Q	2	4	2	1	1	5	2	12	7	3
B6Q6S1	D	3	4	4	2	4	8	3	19	12	6
B6Q7S1	P	2	3	3	2	3	4	4	7	8	7
B6Q8S1	P	4	8	6	5	5	8	5	15	10	11
B6Q9S1	D	3	4	5	3	3	10	7	12	5	6
B6Q10S1	Q	3	7	6	2	3	9	7	14	7	7
B6Q11S1	D	3	6	7	4	4	13	9	20	11	10
B6Q12S1	D	0	0	4	4	0	0	0	1	1	0
B6Q13S1	D	4	12	4	4	6	9	4	16	7	12
B7Q1S1	D	1	1	1	0	1	2	1	5	2	1
B7Q2S1	Q	2	5	3	3	2	6	4	13	8	6
B7Q3S1	D	1	2	2	0	1	3	1	9	4	3
B7Q4S1	D	2	2	3	2	4	5	2	20	3	6
B7Q5S1	Q	6	8	12	6	8	9	9	32	9	15
B7Q6S1	Q	5	10	16	6	8	8	8	32	9	17
B7Q7S1	D	3	5	5	2	3	5	4	18	5	12
B7Q8S1	D	6	12	10	9	6	11	9	31	12	19
B7Q9S1	D	5	6	8	6	4	7	6	26	9	16
B7Q10S1	P	6	8	11	10	6	8	8	21	13	19
B7Q11S1	P	9	11	9	10	6	10	8	31	11	17
B7Q12S1	P	13	7	12	8	8	16	12	41	14	31
B7Q13S1	P	0	0	4	4	0	9	0	1	1	0
B7Q14S1	P	0	1	0	1	0	1	0	3	1	0
B7Q15S1	P	0	1	1	0	1	0	0	3	1	0

Procedures for Handling Missing Data in the U.S. National Adult Literacy Survey

National Adult Literacy Survey participants were asked to answer an extensive set of background questions before they were given the cognitive assessment. Answering these background questions required no reading skills; questions were read to respondents by the interviewers. A Spanish version of the background questionnaire and bilingual interviewers were available to assist individuals whose native language was not English.

As a result of these procedures, a great deal is known about individuals who answered the background questions, including their country of birth, language(s) spoken or read, age, highest level of education completed, current educational aspirations, labor market status, current occupation and wages, voting behaviors, and reading habits. Respondents who did not answer a sufficient number of background questions were considered incomplete cases, and these cases were never incorporated into the database.

After answering the background questions, respondents were given the cognitive assessment. There are three types of non-response with respect to the cognitive items: not presented, omitted, and not attempted/not reached.

- **Not presented**

  Due to the matrix sampling design used in the NALS, each examinee was given only a fraction of the literacy tasks in the item pool. Accordingly, some items were "not presented" to each examinee. Appropriately, these missing data were ignored when calculating respondents' proficiencies.

- **Omitted**

  In some cases, responses were missing because examinees chose not to perform a task or tasks presented to them, either because they were unable to do so or for some other reason. Any missing response(s) that were followed by a valid response (whether correct or incorrect) are defined as "omitted" responses. Omitted responses were treated as wrong, since a random response to an open-ended item would almost certainly result in a wrong answer.

- **Not attempted/not reached**

  In some cases, respondents were given tasks that they did not attempt or did not reach during the time period allotted for the survey. Consecutive missing responses at the end of a block are termed "not attempted/not reached."

The third category is of the greatest concern. A total of 1,364 (5 percent) of the NALS participants did not attempt to complete any of the cognitive items, and another 1,630 (6 percent) completed fewer than five items on each literacy scale, the minimum number needed to estimate their proficiencies accurately. In these cases, the interviewer used a Non-interview Report (NIR) form to document the extent to which the background questions and cognitive tasks were answered and to ascertain the reason for nonresponse. These reasons included:

- native language other than English (specify language)
- physical or mental disability (indicate specific disability: learning disability, mental or emotional condition, mental retardation, hearing impairment, visual impairment, speech or language impairment, physical disability, or other)
- reading and/or writing difficulty
- refused to participate
- other individual refused for respondent
- interviewer made maximum number of attempts to contact sampled person, but no interview was conducted
- sampled person was unavailable during the survey period
- other reason

Since interviewers were obviously not diagnosticians, information about disabilities was recorded by the interviewers as reported either by non-respondents or by those assisting them.

Answers to the NIR questions were subsequently used to create a variable which categorized individuals' reasons for non-response into two groups: *reasons related to literacy* and *reasons unrelated to literacy*. Each of these categories contained several subcategories. As shown in Table L.3, about 12 percent of the NALS participants completed fewer than five cognitive items. Roughly equivalent percentages of these non-respondents cited reasons related to literacy and reasons unrelated to literacy. These reasons are discussed below.

Table L.3: **Reasons for non-response: NALS**

Number of cognitive items completed/ Reason for stopping assessment			Percentage of adults
Completed five or more items			87.8
Completed fewer than five items			
	Stopped	for reasons related to literacy	
		Non-English language, reading difficulty	5.1
		Mental retardation, learning disability	0.6
	Stopped	for reasons not related to literacy	
		Refused	2.7
		Physical disability	2.0
		Other, missing information	1.9
Total			100.0

Reasons related to literacy

Some individuals (5.1 percent of the total sample) did not complete the assessment because they had difficulty reading in the English language. In most cases, this is because English was their second language. The assessment was not designed to investigate the nature of language difficulties in detail. Rather, it was designed to identify language or reading difficulties not caused by physical disabilities. Accordingly, individuals in this category do not necessarily have difficulty reading or writing in languages other than English.

The general household population is likely to be more diverse than the population of individuals in educational institutions. For example, mentally disabled or learning disabled populations are seldom included in the sampling frameworks of traditional school-based assessments. The NALS included every household member within a specified age range, however. Less than 1 percent (0.6) of the total sample did not respond to the survey because they identified themselves (or other identified them) as having a mental or learning disability. Studies of the mentally disabled population have reported that about 1 to 2 percent of the total population is mentally disabled. Such disabilities may be present at birth or may be related to aging or other factors.

Reasons unrelated to literacy

To answer the cognitive items, respondents had to be able to read materials and respond to tasks in writing without help. Individuals with physical disabilities such as visual impairment or lack of motor skills (2 percent of the total sample) would not have been able to respond because they were physically unable to produce written responses. Their reason for not responding was considered unrelated to literacy skills because they may have been able to respond successfully if the modes of presentation and/or response were not paper and pencil.

Those who developed the NALS sought to keep the time burden to a minimum; they also presented easy items at the beginning of the cognitive assessment booklets in order to encourage respondents to continue. (In fact, nearly 98 percent of those who attempted the easy items answered them correctly.) Despite these efforts, some individuals (2.7 percent) refused to complete the cognitive assessment. For subjects who refused to continue after the background questions, no information is available about their performance on the cognitive items. The only information available about their literacy skills is from their answers to certain background questions, such as the highest level of education attained.

About 1.9 percent of the total sample did not respond to the cognitive items because the interviewer exceeded the maximum number of contacts without making an appointment, or because the respondent was unavailable, or for other unspecified reasons. Further, in some cases, individuals were placed in this "other" category because the reasons why they did not complete the survey were unknown. In other words, some individuals in this category might actually belong in one of following categories denoting literacy-related reasons for stopping the assessment, but there was not enough information to make such determinations.

Imputation Methods Considered

Five imputation methods were considered as ways to represent the range of implied causes of missing responses in the NALS. Each of the five methods has a specific effect on the likelihood function. This is important because the posterior distribution is the product of the prior distribution and the likelihood function. Regardless of which of the five methods is used, the prior distribution is the same. The posterior distribution could be different, however, due to different likelihood functions. The five methods differ with respect to the assumed cause of consecutively missing responses at the end of the test for those cases where the respondent answered fewer than five items on at least one scale.

The five imputation methods considered were:

1. Treat all consecutively missing responses of any subject as wrong responses.
2. Treat all consecutively missing responses as wrong if the reason for not responding to the literacy tasks is related to literacy (e.g., language difficulty, mental disability, or reading difficulty not caused by physical disability). Treat the consecutively missing responses of nonrespondents citing reasons unrelated to literacy (e.g., physical disability or refusal) as not reached.
3. For individuals who indicate literacy-related reasons for not responding to the tasks and whose consecutively missing responses start from one of the first five items, treat the consecutively missing responses to the first five items as wrong responses, and treat the remaining portion of the consecutively missing responses as not reached. Treat the consecutively missing responses of individuals who cite a reason not related to literacy as not reached.
4. Regardless of the reasons for nonresponse, if the consecutively missing responses start within the first five items, treat the consecutively missing responses within the first five items as wrong responses, and treat consecutively missing responses after the fifth item as not reached, as is the case in method 3.
5. Treat all consecutively missing responses as not reached.

Table L.4 summarizes the score assignments for each of the five methods. A score of 0 indicates wrong, and a score of 3 indicates not reached.

Table L.4: Treatment of consecutively missing items in the five imputation methods

Method	Reason for nonresponse	Consecutively missing responses for first five items	Remaining consecutively missing items
1	Not related to literacy	0	0
	Related to literacy	0	0
2	Not related to literacy	3	3
	Related to literacy	0	0
3	Not related to literacy	3	3
	Related to literacy	0	3
4	Not related to literacy	0	3
	Related to literacy	0	3
5	Not related to literacy	3	3
	Related to literacy	3	3

It should be noted that Methods 2 and 3 take into account the reason for nonresponse (related or unrelated to literacy), while Methods 1, 4, and 5 do not. Methods 1 and 5 are extremes. According to Method 1, physical disability and refusal are considered synonymous with inability to read. Most large scale assessments use Method 5. This method may not be so extreme if the survey population is fairly uniform—for example, an in-school population. The adult population as a whole is not uniform, however. Because it includes non-English speakers who reside in the United States, some of whom do not read English at all, method 5 ignores too much information. Moreover, it is unlikely that not attempted items occur because of speededness—that is, because the respondent was not given enough time to answer them. The survey was not timed; hence, it is not a speeded test. There were, of course, some practical time limits so that the interviewer would not have to spend an unreasonable number of hours collecting information. On the other hand, respondents were encouraged to attempt all questions.

It was expected that including the nonresponding individuals with individuals who responded to more than five items per scale would result in somewhat lower overall proficiency means. For groups whose characteristics are not related to reasons for nonresponse (e.g., gender), mean proficiencies are affected by all five methods equally. However, variables that do interact with reasons for nonresponse (e.g., level of education) could produce different results depending on which of the five methods is used.

Table L.5 shows the result of these analyses for the prose scale. The prose proficiencies presented here were standardized to have a mean of 50 and a standard deviation of 10 based on those who responded to more than five items.

Table L.5: **Consequences of five methods for selected subpopulation means, prose scale**

	Completed 5+ items		Completed <5 items						Total (both groups)					
	N	Mean	N	1	2	3	4	5	N	1	2	3	4	5
Total	**1,707**	**50**	**154**	**29**	**38**	**39**	**33**	**43**	**1,861**	**48**	**49**	**49**	**49**	**50**
Sex														
Male	775	50	63	29	38	40	33	44	838	49	49	49	49	50
Female	917	50	88	29	37	39	32	43	1,005	48	49	49	48	49
Race/Ethnicity														
White	1,370	51	95	30	42	43	34	45	1,465	50	51	51	50	51
Black	171	42	18	27	35	35	29	38	189	41	42	42	41	42
Hispanic	124	46	32	28	30	32	31	40	156	42	43	43	43	45
Asian	26	50	8	30	29	33	33	45	34	45	45	46	46	48
Am. Indian	11	49	0	0	0	0	0	0	11	49	49	49	49	49
Other	3	46	1	30	30	34	34	47	4	42	42	43	43	46
Education														
Still in high school	68	48	1	30	30	34	34	46	69	48	48	48	48	48
0 to 8 years	94	38	49	27	30	32	29	38	143	34	35	36	35	38
9 to 12 years	180	42	25	28	37	38	31	40	205	41	42	42	41	42
GED	54	46	1	26	34	34	29	39	55	46	46	46	46	46
H.s. diploma	405	48	30	30	45	45	34	46	435	47	48	48	47	48
Some post-secondary	438	52	18	32	43	45	37	51	456	51	52	52	52	52
2 or 4 year degree	437	57	18	33	49	50	38	54	455	56	56	56	56	57
Don't know	3	35	4	29	32	34	32	42	7	31	33	35	33	39
None in U.S.	18	41	8	27	27	30	30	38	26	37	37	38	38	40
Age														
16 to 20	156	49	6	28	28	32	32	42	162	48	48	48	48	48
21 to 25	153	52	10	30	39	41	34	45	163	50	51	51	51	51
26 to 31	211	53	9	28	32	35	32	42	220	52	52	52	52	53
32 to 45	544	52	29	30	37	40	34	46	573	51	51	51	51	52
46 to 64	356	49	38	30	40	41	34	46	394	48	49	49	48	49
65 and older	287	44	62	28	38	39	31	41	349	41	43	43	42	44
Income														
< $5,000	61	46	7	28	35	36	31	41	68	44	45	45	45	46
$5-9,999	98	45	27	28	34	35	31	40	125	41	42	43	42	44
10-14,999	142	46	17	28	34	36	30	40	159	44	44	45	44	45
15-19,999	128	48	14	29	33	35	32	42	142	46	46	47	46	47
20-29,999	241	49	17	28	39	40	32	42	258	48	48	48	48	48
30-39,999	237	51	8	29	36	38	32	43	245	50	50	50	50	51
40-49,999	204	51	9	29	37	39	33	44	213	50	51	51	50	51
$50,000+	465	55	16	33	45	47	38	52	481	54	54	54	54	55
Refused	40	46	24	31	46	46	35	47	64	40	46	46	42	46
Don't know	64	45	14	28	31	34	31	40	78	42	42	43	42	44

The impact of the five methods on the grand total means by gender was nearly zero. Among less educated respondents, younger respondents, and respondents with smaller household income, however, proficiency means varied with all five methods. Not all minority populations were affected in the same way by each of the five methods. For the Black population, the variation was about the same as that for the White population (10 percent of a standard deviation), and there was very little variation according to which method was used. The Hispanic and Asian populations showed the strongest sensitivity to the five methods (about 30 percent of a standard deviation). This is because many of the nonresponding Hispanic and Asian persons indicated a literacy-related reason for not completing the assessment. It is clear that excluding the nonresponse samples would severely overestimate the literacy proficiencies of the Hispanic and Asian populations. To some degree, the proficiency means of less educated, younger, and poorer subpopulations would also be overestimated.

Imputation Method Selected

Methods 2, 3, and 4 have a relatively invariant impact on mean proficiency. Further, Method 2 does not treat nonresponses differentially based on an arbitrary minimum of five items. Accordingly, Method 2 was selected as the most viable approach for including individuals without cognitive data into the sample. To reiterate, the Method 2 imputation procedure relies on background variables and self-reported reasons for nonresponse in addition to the functional relationship between background variables and proficiency scores for the total population.

Using this imputation procedure, all proficiency values were obtained based on two types of information: responses to the background questions and responses to the cognitive items. As an intermediate step, a functional relationship between these two sets of information was calculated for the total sample, and this function was used to obtain unbiased proficiency estimates with reduced error variance. A respondent's proficiency is calculated from a posterior distribution that is the multiple of two functions: the conditional distribution of proficiency, given the pattern of background variables, and the likelihood function of proficiency, given the pattern of responses to the cognitive items.

Since exact matches of background responses are quite rare, NALS used more than 200 principal components to summarize the background information, capturing more than 99 percent of the variance. This procedure was selected to minimize the bias that could result from the inclusion of nonresponding individuals. It implies that the relationships between background variables and proficiency are the same for responding and nonresponding individuals. With this assumption, not-reached coding has no bearing on the proficiency estimates given identical background information. In other words, the higher rate of refusals within certain subpopulations does not lower the mean proficiencies of those subpopulations, unless their responses to the background questions include a high percentage of false statements. Any subpopulation's heterogeneity is evidenced by the wide distribution of background variables. Samples that received a not-reached coding may reflect a particular distribution within a subpopulation. In such cases, the proficiency distribution for nonrespondents would differ from the distribution of respondents because of background variable differences alone.

It is important to reiterate that Method 2 relies on individuals' self-reported reasons for nonresponse. The validity of such reasons is an important concern, given that the method would fail without assurance from validity checks. Because respondents' anonymity and privacy had to be protected in this survey, there was no way to go back to collect further information about nonrespondents' reasons for not completing the assessment. Only internal validity checks were possible, and these supported the validity of respondents' reasons for nonresponse. More detailed information can be found in the *Technical Report on the National Adult Literacy Survey*.

BEST COPY AVAILABLE

Appendix M:

Additional Documentation

Chapter 12

Table M.1: **Parameter coefficients from the logistic regressions, by country**

	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
Prose										
Level 1										
INTER.	4.23	7.81	28.00	9.50	8.96	2.62	11.82	(1.37)	(1.83)	2.96
SEX	0.32	1.26	0.16	0.09	0.59	0.86	0.09	(0.09)	0.59	0.51
AGE	(0.27)	0.34	(1.38)	(0.36)	(0.49)	(0.32)	(0.23)	(0.15)	0.31	(0.17)
AGE2	(0.01)	(0.09)	0.25	0.01	0.05	0.08	(0.02)	(0.03)	(0.07)	0.01
IMM	(3.48)	(2.96)	(2.38)	(0.96)	(1.75)	(2.97)	(3.70)		(2.92)	(1.70)
IMYR	0.08	(0.13)	(0.74)	(0.22)	(0.18)	(0.03)	(0.06)		0.09	
IMYR2	(0.00)	0.00	0.00	0.00	0.00	0.00	(0.00)		(0.00)	
INC	0.11	(0.01)	0.96	0.31	(0.73)	0.29	(0.24)	0.11	0.30	(0.53)
INC2	(0.01)	0.01	(0.12)	(0.05)	0.14	0.07	0.05	(0.02)	(0.02)	0.12
LFS1	1.25	0.61	0.67	0.33	0.86	1.03	0.52	(0.30)	(0.15)	(0.32)
LFS2	1.19		0.24		0.42	1.41	(0.45)			(1.66)
LFS3	0.15	(0.48)	1.96	(0.10)	(0.25)	1.26	0.38	(0.29)	(0.32)	(0.28)
LFS4	2.76	0.56	1.57	0.51	0.18	0.97	0.58	0.24	1.78	0.81
LFS5	1.43		1.08		(0.65)	0.86	0.32			(0.33)
EDUC	1.64	1.56	1.11	0.69	1.03	0.94	1.51	1.35	1.98	0.78
EDUC2	(0.13)	(0.11)	(0.06)	(0.03)	(0.08)	(0.05)	(0.13)	(0.10)	(0.18)	(0.01)
TV	(0.18)	(0.16)	(0.06)	0.04	(0.14)	(0.17)	(0.34)		0.09	
M_ED	(0.14)	(0.08)	0.14	(0.09)	0.59	0.20	0.08		(0.04)	(0.00)
F_ED	(0.01)	(0.02)	0.06	0.01	(0.01)	0.13	0.12		0.01	(0.00)
Level 2										
INTER.	2.08	22.14	(4.34)	5.57	3.32	(0.31)	(0.93)	1.37	(1.95)	0.91
SEX	0.35	(0.46)	(0.14)	(0.00)	(0.20)	0.28	0.34	0.12	(0.18)	0.13
AGE	(0.06)	0.21	0.28	(0.26)	0.37	0.50	(0.49)	(0.30)	(0.10)	(0.17)
AGE2	0.01	0.00	(0.11)	0.03	(0.11)	(0.11)	0.05	0.10	0.02	0.01
IMM	(0.66)	10.29	1.04	(0.35)	0.71	0.91	(1.16)		1.42	0.05
IMYR	0.02	(1.32)	0.10	(0.07)	(0.13)	(0.04)	0.12		0.04	
IMYR2	(0.00)	0.01	(0.00)	0.00	0.00	0.00	(0.00)		(0.00)	
INC	0.09	0.26	0.11	(0.01)	0.09	0.29	0.18	0.07	(0.03)	(0.62)
INC2	(0.01)	(0.03)	(0.01)	0.00	(0.02)	(0.04)	(0.03)	(0.01)	(0.00)	0.11
LFS1	0.44	(0.31)	(0.20)	0.07	(0.91)	(0.10)	(0.26)	0.51	(0.05)	0.02
LFS2	0.47		0.40		(0.93)	(0.30)	0.14			0.20
LFS3	0.90	0.44	(0.37)	0.18	(0.81)	0.13	(0.41)	0.24	(0.04)	0.02
LFS4	0.54	1.36	0.84	(0.29)	(0.08)	0.64	(0.13)	1.45	(0.05)	0.77
LFS5	0.69		(0.17)		(0.80)	0.84	(0.41)			0.32
EDUC	(0.61)	(0.15)	(0.40)	(0.23)	(0.01)	(0.82)	0.36	(1.05)	(1.18)	0.37
EDUC2	0.10	0.04	0.06	0.06	0.03	0.12	(0.01)	0.14	0.16	0.00
TV	(0.24)	0.05	(0.08)	(0.18)	0.00	(0.19)	0.01		(0.06)	
M_ED	0.07	(0.02)	0.08	0.10	0.13	0.04	0.00		0.06	(0.00)
F_ED	0.00	0.09	0.10	0.01	0.06	0.06	0.09		(0.02)	0.01
Level 3										
INTER.	(0.98)	(5.38)	(5.61)	4.02	(11.90)	2.22	(2.44)	4.19	15.57	1.86
SEX	0.28	0.39	0.18	(0.02)	(0.01)	(0.19)	(0.19)	0.07	(0.18)	(0.04)
AGE	0.10	(0.55)	0.16	0.13	0.14	0.17	0.13	0.52	(0.20)	(0.05)
AGE2	(0.01)	0.10	0.01	0.01	(0.01)	(0.04)	0.02	(0.10)	0.08	(0.00)
IMM	1.49	2.73	0.78	0.56	(0.13)	0.73	2.12		1.83	(0.12)
IMYR	0.01	0.01	0.24	(0.07)	0.39	0.01	(0.04)		(0.37)	
IMYR2	0.00	0.00	(0.00)	0.00	(0.00)	(0.00)	0.00		0.00	
INC	0.36	0.26	(0.34)	(0.16)	0.12	(0.42)	0.00	(0.22)	(0.29)	0.11
INC2	(0.05)	(0.03)	0.04	0.02	(0.02)	0.07	(0.01)	0.03	0.03	(0.02)
LFS1	(0.50)	(0.33)	0.08	(0.01)	0.45	(0.72)	(0.10)	(0.39)	0.29	0.03
LFS2	(0.16)		(0.09)		0.79	(0.86)	0.06			0.56
LFS3	(0.61)	(0.10)	(0.26)	(0.10)	0.88	(1.24)	(0.12)	(0.01)	0.37	0.14
LFS4	(0.66)	(0.53)	(0.43)	0.02	0.49	(1.30)	(0.07)	(1.22)	(0.68)	0.18

Table M.1: **Parameter coefficients from the logistic regressions, by country** (Continued)

	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
LFS5	(0.24)		(0.10)		0.89	(1.27)	0.10			0.11
EDUC	(1.48)	(1.52)	(1.08)	(1.15)	(0.35)	(0.94)	(1.13)	(1.44)	(2.12)	(0.87)
EDUC2	0.17	0.17	0.10	0.11	0.04	0.11	0.12	0.13	0.18	0.12
TV	0.01	0.03	0.05	0.11	0.08	0.02	0.01		(0.02)	
M_ED	(0.01)	0.04	(0.07)	(0.00)	(0.07)	(0.12)	0.02		0.02	0.00
F_ED	0.08	0.01	(0.09)	0.02	(0.07)	0.03	(0.06)		0.04	(0.00)
Level 4/5										
INTER.	10.04	12.52	519.34	15.78)	11.91	0.21	0.18	6.18	164.73	0.86
SEX	(0.82)	(1.09)	(0.27)	0.04	0.15	(0.43)	(0.11)	(0.14)	(0.53)	(0.19)
AGE	(0.48)	(0.68)	(0.27)	0.06	(0.86)	(0.71)	0.01	0.80	(0.13)	0.11
AGE2	0.18	0.22	0.12	0.07	0.30	0.15	0.08	0.01	0.02	0.02
IMM	2.95	(3.22)	(0.32)	0.47	1.71	1.89	2.55		178.68	1.09
IMYR	(0.33)	0.39	(11.01)	0.49	(0.21)	0.04	(0.06)		(17.10)	
IMYR2	0.00	(0.00)	0.06	(0.00)	0.00	(0.00)	0.00		0.14	
INC	(0.45)	(0.93)	(0.55)	0.03	0.23	(0.17)	(0.08)	(0.63)	(0.16)	0.57
INC2	0.05	0.12	0.07	0.02	(0.03)	0.00	0.02	0.08	0.03	(0.10)
LFS1	(2.49)	0.95	(0.55)	(0.89)	(0.16)	(0.45)	0.15	(0.29)	(0.34)	0.00
LFS2	(3.16)		0.27		0.18	(0.12)	0.62			0.31
LFS3	(1.89)	0.73	(1.38)	(0.37)	0.31	0.16	0.77	0.20	0.21	(0.00)
LFS4	(3.42)	(1.66)	(1.97)	(0.48)	(0.54)	(0.78)	(0.35)	(1.43)	(0.81)	(0.68)
LFS5	(3.29)		(0.63)		0.28	(0.60)	0.13			(0.12)
EDUC	(0.61)	(1.36)	(0.07)	(0.13)	(0.91)	(0.35)	(1.22)	(0.44)	(2.31)	(0.69)
EDUC2	0.02	0.05	(0.03)	(0.04)	0.04	(0.01)	0.09	(0.05)	0.17	0.03
TV	0.41	(0.14)	0.19	0.05	(0.08)	0.29	0.16		0.03	
M_ED	0.11	0.04	(0.15)	0.01	(0.12)	0.04	(0.08)		(0.11)	0.00
F_ED	(0.14)	(0.23)	(0.17)	(0.10)	0.02	(0.18)	(0.05)		(0.11)	(0.00)
Document										
Level 1										
INTER.	3.62	7.28	29.52	13.42	21.56	0.25	5.91	(1.57)	3.61	3.49
SEX	(0.19)	0.49	(0.36)	0.15	(0.23)	0.69	(0.20)	(0.29)	0.22	0.01
AGE	(0.22)	0.07	(1.15)	(0.34)	(0.55)	(0.13)	(0.43)	(0.12)	0.20	(0.01)
AGE2	(0.01)	(0.06)	0.22	0.01	0.04	(0.00)	0.02	(0.02)	(0.05)	(0.02)
IMM	(2.20)	(1.95)	(2.08)	(1.21)	(1.53)	(2.57)	(3.27)		(2.12)	(1.66)
IMYR	(0.02)	(0.11)	(0.84)	(0.32)	(0.44)	0.02	0.08		(0.07)	
IMYR2	(0.00)	0.00	0.01	0.00	0.00	(0.00)	(0.00)		0.00	
INC	0.36	(0.26)	0.59	0.66	(0.70)	0.44	0.02	0.03	0.32	(0.49)
INC2	(0.04)	0.09	(0.07)	(0.08)	0.11	(0.01)	0.02	(0.01)	(0.04)	0.12
LFS1	1.72	(0.07)	1.46	0.41	0.64	1.04	0.26	0.20	0.06	(0.84)
LFS2	1.65		2.00		(0.22)	1.87	(0.29)			(1.44)
LFS3	1.23	0.10	1.81	1.11	(0.80)	1.20	0.32	(0.03)	(0.14)	(0.77)
LFS4	3.56	0.43	2.66	2.77	(1.34)	1.33	0.86	1.56	1.72	0.34
LFS5	1.62		1.53		(0.79)	1.00	(0.01)			(0.59)
EDUC	1.76	1.72	1.53	0.90	0.84	1.05	1.42	1.31	1.97	1.13
EDUC2	(0.16)	(0.16)	(0.11)	(0.05)	(0.07)	(0.07)	(0.12)	(0.10)	(0.18)	(0.06)
TV	(0.17)	(0.37)	0.03	0.02	(0.15)	(0.11)	(0.29)		0.04	
M_ED	(0.11)	0.04	0.15	(0.01)	0.82	0.10	0.13		(0.02)	0.00
F_ED	(0.04)	(0.03)	0.04	(0.10)	0.16	0.14	0.09		(0.00)	(0.00)
Level 2										
INTER.	3.51	0.81	(3.36)	8.33	(11.69)	1.04	4.87	1.73	(4.86)	1.68
SEX	0.05	0.22	(0.18)	(0.02)	0.10	(0.00)	(0.13)	0.08	(0.15)	(0.24)
AGE	(0.39)	0.21	(0.15)	(0.31)	0.29	0.58	(0.46)	(0.24)	(0.09)	(0.25)
AGE2	0.04	(0.01)	(0.02)	0.03	(0.10)	(0.14)	0.05	0.07	0.03	0.01
IMM	(1.21)	(1.07)	0.37	(0.18)	0.10	0.62	(0.31)		0.47	(0.23)
IMYR	0.09	0.16	0.14	(0.18)	0.27	(0.04)	(0.09)		0.15	
IMYR2	(0.00)	(0.00)	(0.00)	0.00	(0.00)	0.00	0.00		(0.00)	
INC	0.49	(0.19)	0.39	0.05	(0.13)	0.06	0.04	(0.15)	0.10	(0.38)
INC2	(0.06)	(0.00)	(0.04)	(0.01)	0.03	0.01	(0.01)	0.02	(0.01)	0.09
LFS1	(1.11)	0.74	(0.51)	(0.22)	(0.11)	(0.13)	0.37	0.23	(0.19)	0.24
LFS2	(0.31)		0.12		0.33	(0.81)	0.04			0.07
LFS3	(1.40)	0.69	(0.15)	(0.30)	(0.53)	(0.10)	0.01	0.12	0.11	0.22
LFS4	(1.38)	1.66	0.02	(0.60)	0.48	0.23	0.25	0.40	0.38	0.40
LFS5	(0.56)		(0.02)		(0.52)	0.27	0.11			1.02
EDUC	(0.51)	(1.44)	(0.17)	(0.03)	0.71	(0.79)	0.24	(0.62)	(0.95)	0.32
EDUC2	0.08	0.25	0.04	0.05	(0.06)	0.12	(0.01)	0.08	0.14	(0.01)

Table M.1: **Parameter coefficients from the logistic regressions, by country** (Continued)

	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
TV	0.06	(0.13)	(0.03)	(0.01)	(0.11)	(0.08)	(0.04)		0.03	
M_ED	0.15	(0.08)	0.12	0.07	(0.02)	0.06	0.02		0.04	(0.00)
F_ED	(0.01)	0.11	0.03	0.05	0.06	0.09	0.14		(0.03)	0.00
Level 3										
INTER.	0.04	(24.76)	(7.45)	(0.16)	7.53	2.45	(3.69)	3.89	8.48	1.40
SEX	0.04	0.17	0.24	(0.13)	(0.22)	(0.15)	(0.02)	0.23	0.12	(0.01)
AGE	0.08	(0.57)	0.18	(0.12)	0.03	(0.33)	(0.01)	0.43	(0.15)	(0.12)
AGE2	0.01	0.10	(0.02)	0.03	0.00	0.07	0.04	(0.09)	0.04	0.01
IMM	1.89	21.63	0.69	0.35	(0.02)	0.83	0.47		1.65	0.07
IMYR	(0.07)	(0.89)	0.31	0.07	(0.13)	0.02	0.13		(0.18)	
IMYR2	0.00	0.01	(0.00)	(0.00)	0.00	(0.00)	(0.00)		0.00	
INC	(0.30)	0.71	(0.37)	(0.12)	0.26	(0.21)	0.04	0.22	(0.41)	(0.23)
INC2	0.04	(0.08)	0.04	0.01	(0.05)	0.03	(0.01)	(0.02)	0.05	0.03
LFS1	0.27	(0.88)	(1.00)	(0.22)	0.06	(1.13)	(0.53)	(0.48)	0.08	0.28
LFS2	(0.37)		(2.01)		(0.15)	(1.15)	(0.20)			0.69
LFS3	0.43	0.73	(1.39)	(0.31)	0.97	(1.24)	(0.64)	(0.10)	(0.18)	0.22
LFS4	0.03	(1.17)	(1.26)	(0.53)	0.71	(1.64)	(0.35)	(0.87)	(0.59)	0.28
LFS5	(0.41)		(1.56)		0.68	(1.51)	(0.24)			0.15
EDUC	(0.63)	(0.88)	(1.42)	(1.01)	(0.50)	(0.82)	(0.76)	(1.80)	(1.89)	(0.62)
EDUC2	0.06	0.08	0.15	0.10	0.07	0.08	0.08	0.19	0.18	0.10
TV	(0.05)	0.27	0.05	0.02	0.08	0.03	0.05		(0.11)	
M_ED	(0.07)	0.11	(0.04)	(0.06)	(0.01)	(0.15)	0.02		0.01	(0.00)
F_ED	0.13	(0.09)	0.01	0.02	(0.07)	0.00	(0.04)		0.02	0.00
Level 4/5										
INTER.	0.95	22.74	(13.05)	(9.09)	(2.64)	1.63	3.20	4.46	11.86	0.91
SEX	0.03	(0.60)	0.09	0.20	0.47	(0.16)	0.33	0.36	(0.22)	0.20
AGE	0.04	(0.62)	0.08	0.68	(0.46)	(0.13)	0.17	0.13	0.33	0.17
AGE2	0.08	0.22	0.08	(0.07)	0.19	0.06	0.03	0.08	(0.04)	0.02
IMM	0.74	(1.97)	0.96	0.39	1.47	1.67	2.23		1.48	0.84
IMYR	0.04	(0.02)	0.41	0.35	0.07	0.00	(0.18)		(0.18)	
IMYR2	0.00	(0.00)	(0.00)	(0.00)	(0.00)	0.00	0.00		0.00	
INC	(0.12)	(0.91)	(0.30)	(0.40)	0.10	(0.29)	(0.13)	(0.23)	(0.26)	0.66
INC2	0.00	0.11	0.05	0.06	(0.01)	0.01	0.02	0.03	0.02	(0.12)
LFS1	(1.64)	(1.03)	0.24	0.20	(0.38)	(0.09)	0.25	(0.80)	(0.24)	(0.55)
LFS2	(1.54)		0.61		0.43	0.85	1.08			(0.28)
LFS3	(1.03)	(3.05)	(0.25)	(0.05)	(0.15)	0.01	0.91	(0.45)	(0.22)	(0.41)
LFS4	(2.06)	(2.36)	(0.51)	0.38	(0.85)	(0.61)	(0.24)	(1.91)	(1.36)	(0.84)
LFS5	(0.92)		0.53		0.14	0.01	0.51			(0.68)
EDUC	(1.72)	(3.20)	(0.14)	(1.49)	(1.08)	(0.67)	(0.88)	(0.57)	(1.81)	(0.80)
EDUC2	0.17	0.30	(0.03)	0.11	0.08	0.03	0.07	0.01	0.13	0.05
TV	0.14	(0.02)	(0.03)	(0.05)	0.07	0.14	0.11		0.14	
M_ED	0.06	(0.16)	(0.12)	0.07	(0.00)	0.08	(0.08)		(0.07)	0.00
F_ED	(0.15)	(0.07)	(0.14)	(0.07)	(0.02)	(0.18)	(0.11)		0.03	(0.00)
Quantitative										
Level 1										
INTER.	6.61	6.15	26.27	6.82	25.54	1.37	11.09	(1.51)	(0.17)	2.50
SEX	(0.41)	1.26	(0.44)	0.03	(0.26)	0.52	(0.46)	(0.38)	(0.09)	(0.12)
AGE	(0.04)	0.67	(1.35)	(0.56)	(0.50)	0.22	(0.36)	0.10	0.33	(0.14)
AGE2	(0.05)	(0.15)	0.31	0.03	0.08	(0.06)	0.03	(0.06)	(0.05)	0.02
IMM	(2.03)	(1.81)	(1.79)	(1.57)	(0.89)	(2.77)	(2.94)		(2.31)	(1.63)
IMYR	(0.12)	(0.20)	(0.68)	(0.10)	(0.60)	0.00	(0.07)		0.05	
IMYR2	0.00	0.00	0.00	0.00	0.00	(0.00)	0.00		(0.00)	
INC	0.63	0.21	0.82	0.69	(0.34)	0.27	0.09	0.16	0.44	(0.27)
INC2	(0.07)	(0.05)	(0.10)	(0.10)	0.05	0.09	0.00	(0.02)	(0.05)	0.11
LFS1	1.60	1.13	0.56	0.46	0.52	0.72	0.27	0.10	(0.13)	0.03
LFS2	2.45		(0.02)		0.04	2.41	(0.28)			(0.62)
LFS3	0.48	(0.38)	0.72	0.60	(0.36)	0.69	0.18	(0.14)	(0.17)	0.72
LFS4	3.49	0.16	1.66	1.19	(0.65)	1.10	0.74	1.06	1.47	0.89
LFS5	1.94		1.41		(0.28)	0.93	(0.08)			0.62
EDUC	1.89	1.97	1.10	0.89	0.87	1.01	1.35	1.36	1.98	0.96
EDUC2	(0.17)	(0.18)	(0.08)	(0.05)	(0.08)	(0.08)	(0.12)	(0.11)	(0.18)	(0.05)
TV	(0.24)	(0.31)	(0.02)	0.04	(0.26)	(0.14)	(0.38)		(0.04)	
M_ED	(0.11)	0.21	0.08	(0.09)	0.62	0.04	0.14		(0.05)	0.00

Table M.1: **Parameter coefficients from the logistic regressions, by country** (Concluded)

	Canada/ English	Canada/ French	Switzerland/ German	Switzerland/ French	Germany	United States	Netherlands	Poland	France	Sweden
F_ED	0.00	(0.03)	0.05	(0.13)	0.13	0.20	0.11		0.03	(0.00)
Level 2										
INTER	1.71	(2.81)	3.47	12.22	(4.70)	2.74	0.91	1.99	(0.77)	3.33
SEX	(0.15)	(0.28)	(0.33)	(0.36)	(0.12)	(0.33)	(0.32)	(0.10)	(0.00)	(0.40)
AGE	(0.09)	0.06	(0.24)	0.04	0.38	0.37	(0.41)	(0.36)	(0.01)	(0.03)
AGE2	(0.01)	0.01	(0.00)	(0.02)	(0.12)	(0.08)	0.06	0.10	0.01	(0.02)
IMM	(0.42)	1.23	0.25	0.01	(1.12)	0.78	(1.07)		0.23	(0.33)
IMYR	0.09	0.01	(0.10)	(0.31)	0.21	(0.05)	0.09		0.04	
IMYR2	(0.00)	0.00	0.00	0.00	(0.00)	0.00	(0.00)		(0.00)	
INC	0.47	(0.29)	0.25	(0.02)	(0.20)	(0.18)	0.02	(0.10)	0.12	(0.62)
INC2	(0.05)	0.02	(0.03)	0.00	0.04	0.06	(0.01)	0.01	(0.00)	0.11
LFS1	(1.31)	0.07	0.14	(0.01)	(0.23)	0.09	0.22	(0.11)	(0.44)	(0.71)
LFS2	(0.36)		1.20		(0.17)	(0.15)	(0.12)			(0.80)
LFS3	(1.24)	(0.26)	(0.36)	0.18	(0.33)	0.27	(0.08)	(0.02)	(0.37)	(0.72)
LFS4	(1.18)	1.48	0.01	0.49	(0.24)	0.20	(0.07)	0.28	(0.13)	(0.53)
LFS5	(0.08)		0.27		(0.83)	0.21	(0.07)			0.27
EDUC	(0.70)	(1.38)	0.31	0.44	0.71	(0.78)	0.17	(0.48)	(0.47)	0.19
EDUC2	0.13	0.27	(0.01)	(0.02)	(0.06)	0.12	0.02	0.08	0.08	0.01
TV	(0.03)	0.14	(0.12)	(0.04)	(0.08)	(0.19)	0.02		0.02	
M_ED	0.07	(0.10)	0.00	0.09	0.08	0.10	(0.02)		0.09	(0.00)
F_ED	0.03	0.02	0.09	0.04	0.07	0.06	0.13		(0.06)	(0.00)
Level 3										
INTER	(3.07)	37.08	(15.78)	(0.55)	3.59	2.00	(4.75)	2.52	2.80	0.91
SEX	0.21	(0.03)	0.12	(0.04)	(0.16)	(0.17)	0.03	0.41	(0.02)	(0.14)
AGE	(0.19)	(0.35)	0.42	(0.01)	0.13	(0.09)	0.17	0.33	(0.15)	(0.15)
AGE2	0.08	0.07	(0.08)	0.01	(0.02)	0.01	(0.01)	(0.07)	0.01	0.03
IMM	0.89	1.61	0.50	0.07	0.89	1.13	0.77		1.45	0.30
IMYR	0.09	(0.83)	0.55	0.08	(0.10)	(0.02)	0.12		(0.04)	
IMYR2	(0.00)	0.00	(0.00)	(0.00)	0.00	0.00	(0.00)		0.00	
INC	(0.13)	0.57	(0.14)	0.06	0.09	(0.11)	0.02	0.01	(0.41)	(0.31)
INC2	0.01	(0.05)	0.02	(0.00)	(0.02)	0.02	0.00	0.00	0.05	0.04
LFS1	0.20	(0.60)	(1.08)	(0.49)	0.16	(1.05)	(0.21)	0.09	0.35	0.59
LFS2	(0.96)		(1.35)		0.06	(1.42)	0.03			0.68
LFS3	0.79	0.21	(1.30)	(0.53)	0.20	(1.26)	(0.35)	0.10	0.25	0.41
LFS4	(0.41)	(1.50)	(0.96)	(0.51)	0.85	(1.44)	(0.14)	(0.45)	(0.37)	0.55
LFS5	(0.41)		(1.43)		0.64	(1.11)	0.25			0.65
EDUC	(1.49)	(1.18)	(1.46)	(0.78)	(0.98)	(0.88)	(0.67)	(1.41)	(1.74)	(0.50)
EDUC2	0.18	0.11	0.16	0.08	0.14	0.10	0.07	0.14	0.18	0.08
TV	(0.01)	(0.02)	0.07	(0.01)	0.04	(0.03)	0.03		(0.02)	
M_ED	(0.00)	(0.01)	0.10	0.01	(0.03)	(0.04)	0.02		(0.03)	(0.01)
F_ED	0.05	(0.00)	(0.03)	0.02	(0.07)	(0.05)	(0.05)		0.01	0.01
Level 4/5										
INTER	15.46	15.17	(6.17)	(3.89)	(8.11)	(1.16)	13.08	5.44	25.31	0.10
SEX	0.24	(0.21)	0.41	0.43	0.57	0.36	0.60	0.42	0.31	0.48
AGE	(0.21)	(1.93)	(0.09)	0.03	(0.56)	(0.53)	0.07	(0.03)	(0.04)	0.18
AGE2	0.08	0.52	0.08	0.02	0.17	0.14	0.03	0.05	0.04	(0.01)
IMM	1.12	(1.64)	0.47	0.75	1.23	1.42	5.83		3.55	0.61
IMYR	(0.38)	0.04	0.24	0.17	0.21	0.04	(0.74)		(0.73)	
IMYR2	0.00	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	0.01		0.01	
INC	(0.49)	(0.85)	(0.50)	(0.38)	0.20	(0.27)	(0.12)	(0.40)	(0.42)	0.77
INC2	0.06	0.10	0.06	0.05	(0.03)	0.01	0.01	0.04	0.04	(0.14)
LFS1	(1.47)	(1.12)	0.46	0.35	(0.22)	0.09	(0.01)	(0.16)	0.52	(0.43)
LFS2	(1.68)		(0.44)		0.39	(0.11)	0.49			0.11
LFS3	(1.44)	0.23	1.29	0.08	0.32	0.38	0.71	0.28	0.54	(0.36)
LFS4	(2.00)	(1.59)	(0.18)	(0.30)	(0.56)	(0.01)	(0.19)	(1.18)	(0.42)	(0.69)
LFS5	(1.98)		0.43		0.18	(0.17)	0.14			(0.95)
EDUC	(0.78)	(1.59)	(0.03)	(1.32)	(0.53)	(0.42)	(0.85)	(1.37)	(1.68)	(0.53)
EDUC2	0.03	0.08	(0.02)	0.11	0.02	0.00	0.06	0.11	0.13	0.02
TV	0.18	0.07	0.05	0.04	0.10	0.36	0.11		0.08	
M_ED	0.04	(0.06)	(0.09)	0.00	(0.06)	(0.07)	(0.06)		0.01	0.01
F_ED	(0.12)	(0.09)	(0.13)	(0.03)	(0.01)	(0.09)	(0.08)		0.04	(0.00)

Appendix N:

Additional Documentation

Chapter 13

The following are standardized estimates. Significant parameter estimates at 5% level t-test are in bold type. Block (Blon) estimates were not tested by t-tests; see previous x² tests. Parameter estimates are standardized within-group estimates (Jöreskog & Sörbom 1989, pp. 38-40).

Male	B1P1=+0.70*Lit	+0.15*Blo1		+0.69*B1P1&
Female	B1P1=+0.70*Lit	+0.15*Blo1		+0.69*B1P1&
Male	B1P2=**+0.74*Lit**	+0.14*Blo1	**-0.22*Num**	+0.62*B1P2&
Female	B1P2=**+0.74*Lit**	+0.14*Blo1	-0.04*Num	+0.63*B1P2&
Male	B1P3=**+0.71*Lit**	+0.15*Blo1	**+0.12*Num**	+0.68*B1P3&
Female	B1P3=**+0.66*Lit**	+0.14*Blo1	-0.39*Num	+0.63*B1P3&
Male	B1P4=**+0.69*Lit**	+0.15*Blo1		+0.71*B1P4&
Female	B1P4=**+0.69*Lit**	+0.15*Blo1		+0.71*B1P4&
Male	B1P5=**+0.77*Lit**	+0.15*Blo1	+0.24*Num	+0.57*B1P5&
Female	B1P5=**+0.79*Lit**	+0.16*Blo1	+0.05*Num	+0.59*B1P5&
Male	B2P1=**+0.76*Lit**			+0.65*B2P1&
Female	B2P1=**+0.76*Lit**			+0.65*B2P1&
Male	B2P2=**+0.59*Lit**		+0.04*Num	+0.81*B2P2&
Female	B2P2=**+0.59*Lit**		-0.19*Num	+0.80*B2P2&
Male	B2P3=**+0.56*Lit**			+0.83*B2P3&
Female	B2P3=**+0.56*Lit**			+0.83*B2P3&
Male	B2P4=**+0.76*Lit**		**+0.10*Num**	+0.64*B2P4&
Female	B2P4=**+0.76*Lit**		-0.07*Num	+0.64*B2P4&
Male	B2P5=**+0.78*Lit**			+0.62*B2P5&
Female	B2P5=**+0.78*Lit**			+0.62*B2P5&
Male	B3P1=**+0.74*Lit**	+0.12*Blo3	+0.10*Num	+0.65*B3P1&
Female	B3P1=**+0.70*Lit**	+0.29*Blo3	-0.24*Num	+0.62*B3P1&
Male	B3P2=**+0.67*Lit**	+0.15*Blo3	**+0.21*Num**	+0.70*B3P2&
Female	B3P2=**+0.64*Lit**	+0.38*Blo3	+0.09*Num	+0.66*B3P2&
Male	B3P3=**+0.76*Lit**	+0.11*Blo3	-0.10*Num	+0.63*B3P3&
Female	B3P3=**+0.74*Lit**	+0.27*Blo3	-0.09*Num	+0.61*B3P3&
Male	B3P4=**+0.74*Lit**	+0.12*Blo3		+0.66*B3P4&
Female	B3P4=**+0.71*Lit**	+0.31*Blo3		+0.63*B3P4&
Male	B3P5=**+0.76*Lit**	+0.12*Blo3		+0.64*B3P5&
Female	B3P5=**+0.73*Lit**	+0.30*Blo3		+0.62*B3P5&
Male	B4P1=**+0.76*Lit**	+0.21*Blo4	**+0.23*Num**	+0.57*B4P1&
Female	B4P1=**+0.76*Lit**	+0.21*Blo4	+0.25*Num	+0.57*B4P1&
Male	B4P2=**+0.68*Lit**	+0.16*Blo4	**-0.33*Num**	+0.64*B4P2&
Female	B4P2=**+0.67*Lit**	+0.15*Blo4	-0.37*Num	+0.63*B4P2&
Male	B4P3=**+0.74*Lit**	+0.19*Blo4		+0.65*B4P3&
Female	B4P3=**+0.74*Lit**	+0.19*Blo4		+0.65*B4P3&
Male	B4P4=**+0.66*Lit**	+0.20*Blo4	**-0.10*Num**	+0.72*B4P4&
Female	B4P4=**+0.66*Lit**	+0.20*Blo4	-0.13*Num	+0.71*B4P4&
Male	B4P5=**+0.75*Lit**	+0.10*Blo4	**+0.22*Num**	+0.62*B4P5&
Female	B4P5=**+0.74*Lit**	+0.10*Blo4	+0.25*Num	+0.61*B4P5&
Male	B5P1=**+0.45*Lit**	+0.39*Blo5		+0.80*B5P1&
Female	B5P1=**+0.45*Lit**	+0.39*Blo5		+0.80*B5P1&
Male	B5P2=**+0.70*Lit**	+0.10*Blo5	**-0.19*Num**	+0.68*B5P2&
Female	B5P2=**+0.71*Lit**	+0.11*Blo5	-0.03*Num	+0.69*B5P2&
Male	B5P3=**+0.73*Lit**	+0.12*Blo5		+0.68*B5P3&
Female	B5P3=**+0.73*Lit**	+0.12*Blo5		+0.68*B5P3&
Male	B5P4=**+0.64*Lit**	+0.08*Blo5	**+0.20*Num**	+0.74*B5P4&

*U.S. GOVERNMENT PRINTING OFFICE: 1998-436-454/80555

The following are standardized estimates. Significant parameter estimates at 5% level t-test are in bold type. Block (Blon) estimates were not tested by t-tests; see previous x^2 tests. Parameter estimates are standardized within-group estimates (Jöreskog & Sörbom 1989, pp. 38-40). (Concluded)

Female	B5P4=**+0.63*Lit**	+0.08*Blo5	-0.25*Num	+0.73*B5P4&
Male	B5P5=**+0.67*Lit**	+0.11*Blo5	**+0.16*Num**	**+0.71*B5P5&**
Female	B5P5=**+0.63*Lit**	+0.10*Blo5	-0.39*Num	**+0.67*B5P5&**
Male	B6P1=**+0.63*Lit**		**+0.15*Num**	**+0.76*B6P1&**
Female	B6P1=**+0.63*Lit**		+0.08*Num	**+0.77*B6P1&**
Male	B6P2=**+0.68*Lit**		**+0.16*Num**	**+0.71*B6P2&**
Female	B6P2=**+0.69*Lit**		-0.02*Num	**+0.72*B6P2&**
Male	B6P3=**+0.66*Lit**			**+0.76*B6P3&**
Female	B6P3=**+0.66*Lit**			**+0.76*B6P3&**
Male	B6P4=**+0.67*Lit**		**+0.21*Num**	**+0.71*B6P4&**
Female	B6P4=**+0.65*Lit**		-0.30*Num	**+0.70*B6P4&**
Male	B6P5=**+0.60*Lit**			**+0.80*B6P5&**
Female	B6P5=**+0.60*Lit**			**+0.80*B6P5&**
Male	B7P1=**+0.64*Lit**	**+0.29*Blo7**	+0.21*Num	**+0.68*B7P1&**
Female	B7P1=**+0.66*Lit**	+0.28*Blo7	-0.16*Num	**+0.69*B7P1&**
Male	B7P2=**+0.66*Lit**	+0.20*Blo7	+0.10*Num	+0.71*B7P2&
Female	B7P2=**+0.66*Lit**	+0.20*Blo7	-0.18*Num	+0.71*B7P2&
Male	B7P3=**+0.71*Lit**	+0.18*Blo7		+0.68*B7P3&
Female	B7P3=**+0.71*Lit**	+0.18*Blo7		+0.68*B7P3&
Male	B7P4=**+0.71*Lit**	+0.17*Blo7		+0.69*B7P4&
Female	B7P4=**+0.71*Lit**	+0.17*Blo7		+0.69*B7P4&
Male	B7P5=**+0.72*Lit**	+0.23*Blo7		+0.65*B7P5&
Female	B7P5=**+0.72*Lit**	+0.22*Blo7		+0.65*B7P5&

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-049372-2

ISBN 0-16-049372-2
90000
9 780160 493720